GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:14:40 ; Search time 77.2131 Seconds
(without alignments)
46.460 Million cell updates/sec

Title: US-09-823-418-10
Perfect score: 50
Sequence: 1 TRLTRKDGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	50	100.0	10	2	AAY30691	Aay30691 Apo-B100
2	46	92.0	10	2	AAY30690	Aay30690 Apo-B100
3	42	84.0	11	2	AAW57205	Aaw57205 Apo B bin
4	42	84.0	13	2	AAW57207	Aaw57207 Apo B 100
5	42	84.0	15	2	AAW41261	Aaw41261 Apolipopr
6	42	84.0	15	2	AAW96892	Aaw96892 ApoB-100
7	42	84.0	20	6	ABJ37575	Abj37575 Heparin b
8	42	84.0	22	2	AAW57208	Aaw57208 Apo B 100
9	42	84.0	22	2	AAW57209	Aaw57209 Apo B 100
10	42	84.0	34	5	AAE14541	Aae14541 Human apo
11	42	84.0	36	2	AAW96876	Aaw96876 Nucleic a
12	42	84.0	37	2	AAW64587	Aaw64587 Human apo
13	42	84.0	51	2	AAW96845	Aaw96845 Nucleic a
14	42	84.0	343	4	ABB37687	Abb37687 Peptide #
15	42	84.0	343	4	ABG52504	Abg52504 Human liv
16	42	84.0	377	2	AAR72704	Aar72704 Human apo
17	42	84.0	377	2	AAR34031	Aar34031 Sequence
18	42	84.0	2463	8	ADJ57400	Adj57400 Human apo
19	42	84.0	3923	2	AAY31237	Aay31237 Human Apo
20	42	84.0	4536	2	AAW41262	Aaw41262 Apolipopr
21	42	84.0	4536	2	AAW96826	Aaw96826 Amino aci
22	42	84.0	4560	5	AAU98981	Aau98981 Human apo
23	42	84.0	4561	7	ADD48677	Add48677 Human Pro
24	42	84.0	4563	5	AAO15893	Aao15893 Human apo
25	42	84.0	4563	6	ABR40253	Abr40253 Human ali
26	42	84.0	4563	6	ABU79140	Abu79140 Apolipopr
27	42	84.0	4563	7	ADF43408	Adf43408 Apolipopr
28	42	84.0	4563	8	ADH18871	Adh18871 Human apo
29	42	84.0	4563	8	ADH18870	Adh18870 Human apo
30	42	84.0	4563	8	ADO33445	Ado33445 Human apo
31	42	84.0	4563	8	ADO33447	Ado33447 Human apo
32	42	84.0	4590	4	AAU33184	Aau33184 Novel hum
33	39.5	79.0	11	2	AAY30700	Aay30700 Apo-B100
34	38	76.0	10	2	AAY30682	Aay30682 Apo-B100
35	38	76.0	10	2	AAY30687	Aay30687 Apo-B100
36	38	76.0	359	8	ADL90227	Adl90227 Human enz
37	38	76.0	390	7	ADD44981	Add44981 Rat Prote
38	38	76.0	390	7	ADD44985	Add44985 Rat Prote
39	38	76.0	390	7	ADD44987	Add44987 Human Pro
40	38	76.0	390	7	ADD44983	Add44983 Human Pro
41	38	76.0	390	7	ADJ68663	Adj68663 Human hea
42	38	76.0	390	8	ADQ18720	Adq18720 Human sof
43	38	76.0	397	5	AAE27989	Aae27989 Human enz
44	38	76.0	414	7	ADJ68303	Adj68303 Human hea
45	38	76.0	414	7	ADJ68302	Adj68302 Human hea

ALIGNMENTS

RESULT 1
AAY30691
ID AAY30691 standard; peptide; 10 AA.
XX
AC AAY30691;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can

CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 50; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
||||||||||
Db 1 TRLTRKDGLK 10

RESULT 2
AAY30690
ID AAY30690 standard; peptide; 10 AA.
XX
AC AAY30690;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 92.0%; Score 46; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.02;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
||||||:|||
Db 1 TRLTRKEGLK 10

RESULT 3
AAW57205
ID AAW57205 standard; peptide; 11 AA.
XX
AC AAW57205;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B binding site peptide 2.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 12; Page 52; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B binding site
CC peptide which can be used as a component of a non-naturally occurring,
CC receptor-competent low density lipoprotein (LDL) particle of the present
CC invention. The LDL particle comprises at least 1 peptide component that
CC has at least 1 binding site for an apo B protein receptor and at least 1
CC lipophilic substituent. Also described in the invention are peptides
CC containing an apo B binding sequence with at least 70% identity with
CC sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their dimers. Non-
CC naturally occurring, receptor-competent LDL particles are useful as: (i)
CC drug-targeting vectors for delivering anticancer drugs to cancer cells
CC that express an apo B protein receptor, and (ii) additives for cell
CC culture media especially as growth supplements. Non-naturally occurring,
CC receptor-competent LDL particles do not require the complete apo B
CC sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 11 AA;

Query Match 84.0%; Score 42; DB 2; Length 11;
Best Local Similarity 90.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
|||||| |||
Db 2 TRLTRKRGLK 11

```
RESULT 4
AAW57207
ID   AAW57207 standard; peptide; 13 AA.
XX
AC   AAW57207;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide B.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 13 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 13;
  Best Local Similarity   90.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db          3 TRLTRKRGLK 12


RESULT 5
AAW41261
ID   AAW41261 standard; peptide; 15 AA.
XX
AC   AAW41261;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Apolipoprotein B-100 fragment.
XX
KW   Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW   thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
KW   angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW   prothrombinase complex.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9743311-A1.
XX
PD   20-NOV-1997.
XX
PF   09-MAY-1997;   97WO-GB001255.
XX
PR   09-MAY-1996;   96GB-00009702.
XX
PA   (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI   Bruckdorfer KR,  Ettelaie C;
XX
DR   WPI; 1998-008798/01.
XX
PT   Peptide fragments of apo:lipo:protein B-100 with anticoagulant activity -
PT   used for treating or preventing coagulation, inhibiting angiogenesis,
PT   cell differentiation and apoptosis.
XX
PS   Disclosure; Page 22; 60pp; English.
XX
CC   This sequence is an example of the peptide of the invention. It has the
CC   formula (I), or their variants with one or more internal deletions,
CC   insertions or substitutions, while retaining anti-coagulant properties of
CC   apolipoprotein B-100 (apoB-100). Z1-KAQ-X1-KKNKHRHS-X2-T-Z2 (I) X1 = S or
CC   Y; X2 = T or I; Z1 = the N terminus of the peptide, or 1-47 amino acids
CC   (aa); Z2 = the C terminus of the peptide, a terminal amide group or 1-77
CC   aa. Compositions containing the peptide are used for simultaneous,
CC   separate or sequential treatment of cancer, particularly to prevent
CC   metastatic spread. They are also used to inhibit thromboplastin-mediated
CC   processes, specifically to prevent or reduce blood coagulation (e.g.
CC   during or after surgery or in cases of heart attack, stroke etc.) and to
CC   inhibit angiogenesis, cellular differentiation or apoptosis. KRAD-14,
CC   which is active as such or as part of a 98-aa peptide, inhibits
CC   activation of the prothrombinase complex; and prevents activation of
CC   factor VII on the surface of thromboplastin and of platelets by thrombin.
CC   It binds to the residues 58-66 of thromboplastin. Since (I) are much
CC   smaller than apoB-100, they act more quickly
XX
SQ   Sequence 15 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 15;
  Best Local Similarity   90.0%;  Pred. No. 0.2;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db          1 TRLTRKRGLK 10


RESULT 6
AAW96892
ID   AAW96892 standard; peptide; 15 AA.
XX
AC   AAW96892;
XX
DT   22-APR-1999  (first entry)
XX
DE   ApoB-100 nuclear localisation signal sequence, residues 3353-3367.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
```

KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Claim 19; Fig 13D; 293pp; English.
XX
CC AAW96878-97 represent nuclear localisation signal sequence derived from
CC human apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein
CC component of very-low density lipoproteins (VLDL), intermediate density
CC lipoprotein (IDL), low density lipoproteins ((LDL) and lipoprotein a. The
CC present sequence can be used in the composition of the invention. The
CC specification describes a composition that comprises LDL and
CC apolipoproteins for the binding and in vivo transport of nucleic acids.
CC The composition is used to deliver nucleic acids to eukaryotic cells, in
CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC molecule (or ribozyme). Specifically they are used for gene therapy of
CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC fibrosis and arteriosclerosis
XX
SQ Sequence 15 AA;

Query Match 84.0%; Score 42; DB 2; Length 15;
Best Local Similarity 90.0%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
|||||| |||
Db 6 TRLTRKRGLK 15

RESULT 7
ABJ37575
ID ABJ37575 standard; peptide; 20 AA.
XX
AC ABJ37575;
XX
DT 10-MAY-2003 (first entry)
XX
DE Heparin binding peptide sequence #28.
XX
KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW cardiovascular; circulatory; ligand; sulphated; sulphonated; tumour;
KW rheumatoid arthritis; hypoxia; diabetic retinopathy; heparin.
XX
OS Unidentified.
XX
PN WO2003007689-A2.
XX
PD 30-JAN-2003.
XX
PF 22-JUL-2002; 2002WO-US023419.
XX
PR 20-JUL-2001; 2001US-0306726P.
XX
PA (ETHZ-) ETH ZUERICH.
PA (UYZU-) UNIV ZURICH.
XX
PI Hubbell JA, Schoenmakers R, Maynard HD;
XX
DR WPI; 2003-300420/29.
XX
PT Use of a ligand comprising of at least one sulfated or sulfonated amino
PT acid for the treatment of e.g. tumors, rheumatoid arthritis, diabetic
PT retinopathy and hypoxia.
XX
PS Disclosure; Fig 2; 79pp; English.
XX
CC The invention relates to a novel ligand for binding a target biomolecule,
CC which comprises a peptide having at least one sulphated or sulphonated
CC amino acid and at least one amino acid chosen from neutral and positively
CC charged amino acids. The novel ligands can be used for the treatment of
CC e.g. tumours, rheumatoid arthritis, diabetic retinopathy and hypoxia.
CC This sequence represents a heparin binding peptide relating to the
CC invention
XX
SQ Sequence 20 AA;

Query Match 84.0%; Score 42; DB 6; Length 20;
Best Local Similarity 90.0%; Pred. No. 0.27;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
|||||| |||
Db 7 TRLTRKRGLK 16

RESULT 8
AAW57208
ID AAW57208 standard; peptide; 22 AA.
XX
AC AAW57208;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide C.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
FT Modified-site 22
FT /note= "attached to cholesterol"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.

XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 84.0%; Score 42; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 0.29;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
|||||| |||
Db 7 TRLTRKRGLK 16

RESULT 9
AAW57209
ID AAW57209 standard; peptide; 22 AA.
XX
AC AAW57209;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide D.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 84.0%; Score 42; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 0.29;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
|||||| |||
Db 7 TRLTRKRGLK 16

RESULT 10
AAE14541
ID AAE14541 standard; peptide; 34 AA.
XX
AC AAE14541;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human apoB-100 derived peptide p62.
XX
KW Oxidised low density lipoprotein; oxLDL; immunoassay; periaortitis;
KW cardiovascular disease; coronary heart disease; pre-eclampsia;
KW non-insulin-dependent diabetes; endothelial dysfunction; human; apoB-100;
KW peptide p62.
XX
OS Homo sapiens.
XX
PN WO200206314-A2.
XX
PD 24-JAN-2002.
XX
PF 18-JUL-2001; 2001WO-GB003212.
XX
PR 18-JUL-2000; 2000GB-00017641.
XX
PA (ARKT-) ARK THERAPEUTICS LTD.
XX
PI Narvanen O, Yla-Herttuala S;
XX
DR WPI; 2002-179777/23.
XX
PT New peptide useful in enzyme immunoassays for detecting oxidized low
PT density lipoprotein which is a marker of coronary heart disease and other
PT cardiovascular diseases, has affinity for oxidized low density
PT lipoprotein.
XX
PS Claim 6; Page 5; 21pp; English.
XX
CC The invention relates to peptides having affinity for oxidised low
CC density lipoprotein (oxLDL), in cyclised or multimeric form. The peptide
CC is useful in an immunoassay to determine the presence, and optionally,
CC the amount of antibodies in a sample, having affinity for oxLDL.
CC Preferably immobilised peptide is useful for measuring the amount of
CC autoantibodies for oxLDL in a sample, especially a serum or plasma sample
CC from a patient for evaluating the risk of coronary heart diseases, other
CC cardiovascular diseases, and several other disorders such as
CC periaortitis, pre-eclampsia, non-insulin-dependent diabetes and
CC endothelial dysfunction. The peptide of the invention is stable, can be
CC synthesised easily without the need to isolate proteins from a patient's

```
CC   blood, and has a long half-life. The present sequence is human apoB-100
CC   derived peptide p62 used in the invention
XX
SQ   Sequence 34 AA;

  Query Match             84.0%;  Score 42;  DB 5;  Length 34;
  Best Local Similarity   90.0%;  Pred. No. 0.47;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db         25 TRLTRKRGLK 34


RESULT 11
AAW96876
ID   AAW96876 standard; peptide; 36 AA.
XX
AC   AAW96876;
XX
DT   22-APR-1999  (first entry)
XX
DE   Nucleic acid binding domain from apoB-100, residues 3348-3390.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9856938-A1.
XX
PD   17-DEC-1998.
XX
PF   10-JUN-1998;   98WO-US011927.
XX
PR   13-JUN-1997;   97US-00874807.
PR   14-MAY-1998;   98US-00079030.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Guevara JG,  Hoogeveen RC,  Moore JP;
XX
DR   WPI; 1999-070331/06.
XX
PT   Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT   used for delivering nucleic acid to cells for gene therapy and antisense
PT   treatment.
XX
PS   Claim 16; Fig 12C; 293pp; English.
XX
CC   AAW96827-77 represent nucleic acid binding domains derived from human
CC   apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC   of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC   (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC   sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 36 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 36;
  Best Local Similarity   90.0%;  Pred. No. 0.5;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db         11 TRLTRKRGLK 20


RESULT 12
AAW64587
ID   AAW64587 standard; peptide; 37 AA.
XX
AC   AAW64587;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Human apolipoprotein peptide fragment #1.
XX
KW   Factor V; human; detection; protein function; blood coagulation; apo;
KW   fat metabolism; Leyden mutation; deep vein thrombosis; apolipoprotein;
KW   Alzheimer's disease; 5,10-methylenetetrahydrofolate reductase; prion;
KW   hypercysteinaemia; factor VII; cardiovascular disease; pathogen; virus.
XX
OS   Homo sapiens.
XX
PN   EP857973-A2.
XX
PD   12-AUG-1998.
XX
PF   12-JAN-1998;   98EP-00890007.
XX
PR   13-JAN-1997;   97AT-00000044.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Moritz B,  Kiessig S,  Lang H,  Schenk V;
XX
DR   WPI; 1998-416142/36.
XX
PT   Detecting or quantifying mutant protein in presence of wild-type protein
PT   comprises reaction with ligand - used to detect mutant blood coagulation
PT   factors or apolipoproteins for diagnosing risk of thrombosis.
XX
PS   Example 2; Page 9; 18pp; German.
XX
CC   AAW64587 and AAW64588 are fragments of human apolipoprotein which are
CC   used with Factor V protein fragments in a novel method to detect the
CC   presence of a mutated protein in a sample that may also contain the
CC   corresponding wild-type protein. The method is used to detect mutations
CC   that alter protein functions (either point mutation or small insertions
CC   or deletions), particularly in proteins involved in blood coagulation or
CC   metabolism of fat. Protein functions which are specially detectable are
CC   the Leyden mutation in factor V (associated with increased risk of deep
CC   vein thrombosis), mutations in apolipoprotein (apo) genes (certain
CC   alleles of apoE indicates increased risk of developing Alzheimer's
CC   disease), thermostable 5,10-methylenetetrahydrofolate reductase
CC   (associated with hypercysteinaemia and venous thrombosis) and factor VII
CC   mutations (associated with increased risk of cardiovascular disease). The
CC   method can also be applied to proteins from pathogens, e.g. viruses or
CC   prions. The method does not require complex apparatus for polymerase
CC   chain reactions, it is simple, standardisable and reliable and is
CC   particularly suited to routine screening. It also allows mutant protein
CC   in a sample to be quantified
XX
SQ   Sequence 37 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 37;
  Best Local Similarity   90.0%;  Pred. No. 0.51;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db         11 TRLTRKRGLK 20


RESULT 13
AAW96845
ID   AAW96845 standard; peptide; 51 AA.
```

```
XX
AC   AAW96845;
XX
DT   22-APR-1999  (first entry)
XX
DE   Nucleic acid binding domain from apoB-100.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9856938-A1.
XX
PD   17-DEC-1998.
XX
PF   10-JUN-1998;   98WO-US011927.
XX
PR   13-JUN-1997;   97US-00874807.
PR   14-MAY-1998;   98US-00079030.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Guevara JG,  Hoogeveen RC,  Moore JP;
XX
DR   WPI; 1999-070331/06.
XX
PT   Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT   used for delivering nucleic acid to cells for gene therapy and antisense
PT   treatment.
XX
PS   Claim 16; Page 151; 293pp; English.
XX
CC   AAW96827-77 represent nucleic acid binding domains derived from human
CC   apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC   of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC   (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC   sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 51 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 51;
  Best Local Similarity   90.0%;  Pred. No. 0.72;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db          6 TRLTRKRGLK 15


RESULT 14
ABB37687
ID   ABB37687 standard; peptide; 343 AA.
XX
AC   ABB37687;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #5193 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 27; SEQ ID NO 30322; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a peptide encoded by a single exon nucleic acid probe
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 343 AA;

  Query Match             84.0%;  Score 42;  DB 4;  Length 343;
  Best Local Similarity   90.0%;  Pred. No. 5.5;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db        169 TRLTRKRGLK 178


RESULT 15
ABG52504
ID   ABG52504 standard; peptide; 343 AA.
XX
AC   ABG52504;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human liver peptide, SEQ ID No 31152.
XX
KW   Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW   hypercholesterolaemia; coronary heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
```

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SC, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 27; SEQ ID NO 31152; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 343 AA;

Query Match 84.0%; Score 42; DB 4; Length 343;
Best Local Similarity 90.0%; Pred. No. 5.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TRLTRKDGLK 10
              |||||| |||
Db        169 TRLTRKRGLK 178
```

Search completed: January 13, 2005, 01:43:00
Job time : 79.3798 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:30:05 ; Search time 14.4262 Seconds
(without alignments)
66.696 Million cell updates/sec

Title: US-09-823-418-10
Perfect score: 50
Sequence: 1 TRLTRKDGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	84.0	596	2	S32802	apolipoprotein B -
2	42	84.0	4563	1	LPHUB	apolipoprotein B-1
3	38	76.0	269	2	C60950	apolipoprotein B-1
4	38	76.0	389	1	DEPGPA	pyruvate dehydroge
5	38	76.0	390	1	DEHUPA	pyruvate dehydroge
6	38	76.0	390	1	DERTP1	pyruvate dehydroge
7	38	76.0	390	1	DERTPA	pyruvate dehydroge
8	38	76.0	390	2	S23506	pyruvate dehydroge
9	38	76.0	779	2	JH0102	apolipoprotein B -
10	36	72.0	275	2	E60950	apolipoprotein B-1
11	36	72.0	394	2	T46858	molybdenum cofacto
12	35	70.0	227	2	A84109	ABC transporter (A
13	35	70.0	325	2	T29604	hypothetical prote
14	35	70.0	480	2	G71050	asparagine synthas
15	34	68.0	131	2	AI1540	hypothetical prote
16	34	68.0	190	2	G82634	hypothetical prote
17	34	68.0	857	2	T37459	ribonucleotide red
18	33	66.0	453	2	G83043	probable transport
19	33	66.0	481	2	C82421	conserved hypothet
20	33	66.0	561	1	S34191	sulfite reductase
21	33	66.0	668	1	S74619	hypothetical prote
22	33	66.0	772	2	I50463	protein kinase - c
23	32	64.0	206	2	B97282	ribosomal protein
24	32	64.0	208	2	C82742	transcription regu
25	32	64.0	266	2	C72380	hypothetical prote
26	32	64.0	272	2	E83363	hypothetical prote
27	32	64.0	274	2	A60950	apolipoprotein B-1
28	32	64.0	280	2	S06572	finger protein (cl
29	32	64.0	289	2	A70751	hypothetical prote
30	32	64.0	370	2	A49360	pyruvate dehydroge
31	32	64.0	388	1	DEHUPT	pyruvate dehydroge
32	32	64.0	390	2	AG0929	probable major tai
33	32	64.0	390	2	AI0836	probable bacteriop
34	32	64.0	473	2	H72393	hypothetical prote
35	32	64.0	555	2	C45868	glycerol-3-phospha
36	32	64.0	584	2	D81265	hypothetical prote
37	32	64.0	687	2	S69723	hypothetical prote
38	32	64.0	689	2	AC1408	transcription anti
39	32	64.0	689	2	AC1784	transcription anti
40	32	64.0	784	2	JH0101	apolipoprotein B-1
41	32	64.0	819	2	S43748	translation elonga
42	32	64.0	833	2	T32289	hypothetical prote
43	32	64.0	1199	2	T18348	probable pol polyp
44	31	62.0	26	2	S51055	ribosomal protein
45	31	62.0	151	2	T28840	hypothetical prote

ALIGNMENTS

RESULT 1
S32802
apolipoprotein B - crab-eating macaque (fragment)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S32802
R;Pape, M.E.; Castle, C.K.; Murray, R.W.; Funk, G.M.; Hunt, C.E.; Marotti, K.R.; Melchi
Biochim. Biophys. Acta 1086, 326-334, 1991
A;Title: Apo B metabolism in the cynomolgus monkey: evidence for post-transcriptional r
A;Reference number: S32802; MUID:92075708; PMID:1742325
A;Accession: S32802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-596 <PAP>
A;Cross-references: UNIPROT:Q28473; EMBL:X15737; NID:g38047; PIDN:CAA33755.1; PID:g9301
C;Superfamily: apolipoprotein B

Query Match 84.0%; Score 42; DB 2; Length 596;
Best Local Similarity 90.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TRLTRKDGLK 10
              |||||| |||
Db        226 TRLTRKRGLK 235
```

RESULT 2
LPHUB
apolipoprotein B-100 precursor - human
N;Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
C;Species: Homo sapiens (man)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: A27850; A25679; A25263; A25267; A25266; A24320; A24684; A23817; A25774; A2
4452; I61909; I59510; I39474; I39469; I84624; I37179; PS0058
R;Ludwig, E.H.; Blackhart, B.D.; Pierotti, V.R.; Caiati, L.; Fortier, C.; Knott, T.; Sc
DNA 6, 363-372, 1987
A;Title: DNA sequence of the human apolipoprotein B gene.
A;Reference number: A27850; MUID:88003974; PMID:3652907
A;Accession: A27850
A;Molecule type: DNA
A;Residues: 1-617,'A',619-1929,'F',1931-3318,'D',3320-3426,'T',3428-3431,'Q',3433-3731,
A;Cross-references: UNIPROT:P04114; UNIPROT:P78482; UNIPROT:P78479; UNIPROT:Q9UMN0; UNI
R;Cladaras, C.; Hadzopoulou-Cladaras, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A;Title: The complete sequence and structural analysis of human apolipoprotein B-100: r
A;Reference number: A91058; MUID:87161758; PMID:3030729
A;Accession: A25679
A;Molecule type: mRNA
A;Residues: 1-11,15-2539,'S',2541-3823,'R',3825-4563 <CLA>
A;Note: 1109-Asp was also found
R;Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McC
Nucleic Acids Res. 14, 7501-7503, 1986

A;Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A;Reference number: A93639; MUID:87016385; PMID:3763409
A;Accession: A25263
A;Molecule type: mRNA
A;Residues: 1-272,'N',274-617,'A',619-1217,'E',1219-2091,'V',2093-2364,'T',2366-2679,'Q'
A;Cross-references: GB:X04506; NID:g34330; PIDN:CAA28191.1; PID:g34331
R;Law, S.W.; Grant, S.M.; Higuchi, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer Jr
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A;Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A;Reference number: A94134; MUID:87041416; PMID:3464946
A;Accession: A25267
A;Molecule type: mRNA
A;Residues: 1-617,'A',619-703,'P',705-792,'R',794-1270,'S',1272-1866,'G',1868-2036,'N',2
4189-4220,'M',4222-4563 <LAW>
A;Note: the codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4188, and
R;Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Gotto Jr., A.M
J. Biol. Chem. 261, 12918-12921, 1986
A;Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A;Reference number: A92556; MUID:87008488; PMID:3759943
A;Accession: A25266
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-328,'V',330-644,'I',646-918,'P',920-3318,'D',3320-3426,'T',3428-
9-4132,'G',4134-4180,'E',4182-4563 <CHE>
A;Cross-references: GB:J02610; NID:g178803; PIDN:AAA35549.1; PID:g178804
A;Note: a total of 2366 residues were confirmed by direct sequencing of tryptic peptides
R;Protter, A.A.; Hardman, D.A.; Sato, K.Y.; Schilling, J.W.; Yamanaka, M.; Hort, Y.J.; H
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A;Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein
A;Reference number: A24320; MUID:86287319; PMID:3461454
A;Accession: A24320
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-617,'A',619-941,'YYIWSLPPKP',951-1138,'PTGRLPNCFSNGLICYSLWLHSFQE
A;Cross-references: GB:M14081; NID:g178795; PIDN:AAA51752.1; PID:g553189
R;Law, S.W.; Lackner, K.J.; Hospattankar, A.V.; Anchors, J.M.; Sakaguchi, A.Y.; Naylor,
Proc. Natl. Acad. Sci. U.S.A. 82, 8340-8344, 1985
A;Title: Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of
A;Reference number: A24684; MUID:86094221; PMID:3001697
A;Accession: A24684
A;Molecule type: mRNA
A;Residues: 485-617,'A',619-1044 <LA2>
A;Cross-references: GB:M12480; NID:g178791; PIDN:AAA51751.1; PID:g178792
R;Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; Ki
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A;Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipopr
A;Reference number: A94088; MUID:86149325; PMID:3513177
A;Accession: A23817
A;Molecule type: mRNA
A;Residues: 1-291 <PRO>
A;Cross-references: GB:M12681; NID:g178797; PIDN:AAA51753.1; PID:g178798
R;Deeb, S.S.; Motulsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A;Title: A partial cDNA clone for human apolipoprotein B.
A;Reference number: A25774; MUID:85270450; PMID:3860836
A;Accession: A25774
A;Molecule type: mRNA
A;Residues: 709-791,'SSSWKAASHGCPHSAGD',810-906 <DEE>
A;Cross-references: GB:K03175; NID:g178821; PIDN:AAA51759.1; PID:g178822
R;Carlsson, P.; Darnfors, C.; Olofsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A;Title: Analysis of the human apolipoprotein B gene; complete structure of the B-74 reg
A;Reference number: A91565; MUID:87191999; PMID:2883086
A;Accession: A26533
A;Molecule type: mRNA
A;Residues: 1282-2721,2742-3290,'L',3292-3336,'N',3338-3948,'F',3950-3963,'Y',3965-4180,
A;Cross-references: GB:M15421; NID:g178817; PIDN:AAA51758.1; PID:g178818
R;Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Sato, K.Y.; Lau, K.; Yamana
Biochemistry 26, 5478-5486, 1987
A;Title: Structural comparison of human apolipoproteins B-48 and B-100.
A;Reference number: A29671; MUID:88050832; PMID:3676265
A;Accession: A29671
A;Molecule type: mRNA
A;Residues: 1671-2323,'PYW',2327-2352,'H',2354-2398 <HAR>
A;Cross-references: GB:M17367; NID:g178731; PIDN:AAA51741.1; PID:g178732

R;Shoulders, C.C.; Myant, N.B.; Sidoli, A.; Rodriguez, J.C.; Cortese, C.; Baralle, F.E.
Atherosclerosis 58, 277-289, 1985
A;Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for more than o
A;Reference number: A90084; MUID:86130855; PMID:3841481
A;Accession: A29287
A;Molecule type: mRNA
A;Residues: 3846-4298 <SHO>
R;Pfitzner, R.; Wagener, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986
A;Title: Isolation, expression and characterization of a human apolipoprotein B 100-spe
A;Reference number: A25572; MUID:87076044; PMID:3024665
A;Accession: A25572
A;Molecule type: mRNA
A;Residues: 4219-4337,'S',4339-4563 <PFI>
A;Cross-references: GB:M36676
R;Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow, J.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269, 1985
A;Reference number: A24738; MUID:86042646; PMID:2932736
A;Accession: A24738
A;Molecule type: mRNA
A;Residues: 'N',3729-3731,'I',3733-3875,'A',3877-3948,'F',3950-3963,'Y',3965-3982,'S',3
A;Cross-references: GB:M12413; NID:g178735; PIDN:AAA51742.1; PID:g178736
R;Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.; Silberman, S.R.; Ca
Science 238, 363-366, 1987
A;Title: Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific i
A;Reference number: A40133; MUID:88018019; PMID:3659919
A;Accession: B40133
A;Molecule type: mRNA
A;Residues: 2165-2179 <CH1>
A;Cross-references: GB:M18036; NID:g178799; PIDN:AAA51754.1; PID:g178800
A;Note: this mRNA includes the stop codon of the organ-specific mRNA for apo48
A;Accession: A40133
A;Molecule type: protein
A;Residues: 51-75;101-110;129-139;158-174;197-207;276-287;298-304;306-314;526-532;538-5
36;1486-1498;1537-1556;1563-1572;1601-1610;1647-1661;1697-1724;1770-1781;1859-1897;1968
A;Note: these fragments were derived from apo48
R;Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987
A;Title: Carboxyl terminal analysis of human B-48 protein confirms the novel mechanism
A;Reference number: A28002; MUID:88106542; PMID:3426612
A;Accession: A28002
A;Molecule type: mRNA
A;Residues: 2129-2179,2181-2235 <HA2>
A;Cross-references: GB:M18471
A;Experimental source: intestine
A;Note: this mRNA from intestine includes a stop codon created by RNA editing in place
R;Mehrabian, M.; Schumaker, V.N.; Fareed, G.C.; West, R.; Johnson, D.F.; Kirchgessner,
Nucleic Acids Res. 13, 6937-6953, 1985
A;Title: Human apolipoprotein B: identification of cDNA clones and characterization of
A;Reference number: A24269; MUID:86041888; PMID:3903660
A;Accession: A24269
A;Molecule type: mRNA
A;Residues: 3056-3159 <MEH>
A;Cross-references: GB:X03045; NID:g28783; PIDN:CAA26850.1; PID:g929609
R;Hospattankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer Jr., H.B.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A;Title: Identification of a novel in-frame translational stop codon in human intestine
A;Reference number: A29659; MUID:88049670; PMID:2445342
A;Accession: A29659
A;Molecule type: mRNA
A;Residues: 2169-2179 <HOS>
A;Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A;Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human intest
ch encodes the 250K apoB-48, CAA encoding 2180-Gln is substituted by the stop codon TAA,
R;Yang, C.; Kim, T.W.; Weng, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
A;Title: Isolation and characterization of sulfhydryl and disulfide peptides of human ap
A;Reference number: A35783; MUID:90319144; PMID:2115173
A;Contents: disulfide bonds
A;Accession: A35783
A;Molecule type: protein
A;Residues: 28-41;76-97,'I',99-100;175-193;206-215;239-249;259-266;357-399;455-490;512-9

```
A;Note: cysteines at positions 1112, 1422, 1505, 1662, 3761, 3917, and 4217 have free su
R;LeBoeuf, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
FEBS Lett. 170, 105-108, 1984
A;Title: Human apolipoprotein B: partial amino acid sequence.
A;Reference number: A22006; MUID:84208786; PMID:6373369
A;Accession: A22006
A;Molecule type: protein
A;Residues: 873-892,'K',894-896 <LE1>
A;Accession: B22006
A;Molecule type: protein
A;Residues: 3113,'L',3115-3130,'R',3132-3133,'P',3135-3136,'R' <LE2>
R;Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Caiati, L.; Onasch, M.A.; Wallis, S.C.;
J. Biol. Chem. 261, 15364-15367, 1986
A;Title: Structure of the human apolipoprotein B gene.
A;Reference number: A92564; MUID:87057153; PMID:2946672
A;Contents: annotation; gene structure
R;Wagener, R.; Pfitzner, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
A;Title: Studies on the organization of the human apolipoprotein B 100 gene.
A;Reference number: A90715; MUID:87271140; PMID:2886136
A;Contents: annotation; gene structure
R;Weisgraber, K.H.; Rall Jr., S.C.
J. Biol. Chem. 262, 11097-11103, 1987
A;Title: Human apolipoprotein B-100 heparin-binding sites.
A;Reference number: A92605; MUID:87280197; PMID:3301850
A;Contents: annotation; heparin binding and disulfide bond
R;Dashti, N.; Lee, D.M.; Mok, T.
Biochem. Biophys. Res. Commun. 137, 493-499, 1986
A;Title: Apolipoprotein B is a calcium binding protein.
A;Reference number: A90125; MUID:86242245; PMID:3087360
A;Contents: annotation; calcium binding
R;Carlsson, P.; Olofsson, S.O.; Bondjers, G.; Darnfors, C.; Wiklund, O.; Bjursell, G.
Nucleic Acids Res. 13, 8813-8826, 1985
A;Title: Molecular cloning of human apolipoprotein B cDNA.
A;Reference number: I37178; MUID:86093680; PMID:3841204
A;Accession: I37180

  Query Match             84.0%;  Score 42;  DB 1;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 15;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db       3385 TRLTRKRGLK 3394


RESULT 3
C60950
apolipoprotein B-100 - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: C60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LDL
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: C60950
A;Molecule type: DNA
A;Residues: 1-269 <LAW>
A;Cross-references: UNIPROT:Q60537; UNIPROT:Q60536
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;

  Query Match             76.0%;  Score 38;  DB 2;  Length 269;
  Best Local Similarity   80.0%;  Pred. No. 6.2;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              :||||| |||
Db        216 SRLTRKRGLK 225


RESULT 4
DEPGPA
pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) alpha chain precursor - pig (fragment)
N;Alternate names: pyruvate dehydrogenase complex, E1 component alpha chain
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: S20813
R;Sermon, K.; DeMeirleir, L.; Elpers, I.; Lissens, W.; Liebaers, I.
submitted to the EMBL Data Library, May 1990
A;Reference number: S20813
A;Accession: S20813
A;Molecule type: mRNA
A;Residues: 1-389 <SER>
A;Cross-references: UNIPROT:P29804; EMBL:X52990; NID:g1850; PIDN:CAA37180.1; PID:g1851
A;Experimental source: muscle
C;Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-b
C;Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; phosphoprotein
F;1-28/Domain: transit peptide (mitochondrion) (fragment) #status predicted <TNP>
F;29-389/Product: pyruvate dehydrogenase (lipoamide) alpha chain #status predicted <MA1
F;184-233/Domain: thiamin pyrophosphate-binding domain homology <TPB>
F;231/Binding site: phosphate (Ser) (covalent) #status experimental
F;292/Binding site: phosphate (Ser) (covalent) #status experimental
F;299/Binding site: phosphate (Ser) (covalent) #status experimental

  Query Match             76.0%;  Score 38;  DB 1;  Length 389;
  Best Local Similarity   80.0%;  Pred. No. 8.9;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         53 TVLTREDGLK 62


RESULT 5
DEHUPA
pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) alpha chain precursor - human
N;Alternate names: pyruvate dehydrogenase (PDH) complex, E1 component alpha chain
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1989 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: JQ0770; A33905; A28275; S11715; S11716; A28398; A29577; I56309; I54262; I5
R;Koike, K.; Urata, Y.; Matsuo, S.; Koike, M.
Gene 93, 307-311, 1990
A;Title: Characterization and nucleotide sequence of the gene encoding the human pyruva
A;Reference number: JQ0770; MUID:91033044; PMID:2227443
A;Accession: JQ0770
A;Molecule type: DNA
A;Residues: 1-390 <KOI>
A;Cross-references: UNIPROT:P08559; GB:D90084; NID:g219981; PIDN:BAA14121.1; PID:g21998
A;Experimental source: leukocyte
R;Ho, L.; Wexler, I.D.; Liu, T.C.; Thekkumkara, T.J.; Patel, M.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 5330-5334, 1989
A;Title: Characterization of cDNAs encoding human pyruvate dehydrogenase alpha-subunit.
A;Reference number: A33905; MUID:89315791; PMID:2748588
A;Accession: A33905
A;Molecule type: mRNA
A;Residues: 1-390 <HOL>
A;Cross-references: GB:M24848; NID:g190761; PIDN:AAA36533.1; PID:g190762
R;De Meirleir, L.; MacKay, N.; Wah, A.M.L.H.; Robinson, B.H.
J. Biol. Chem. 263, 1991-1995, 1988
A;Title: Isolation of a full-length complementary DNA coding for human E-1-alpha subuni
A;Reference number: A28275; MUID:88115327; PMID:2828359
A;Accession: A28275
A;Molecule type: mRNA
A;Residues: 1-348,'P',350-353,'A',355-390 <DEM1>
A;Cross-references: EMBL:J03503; NID:g189765
A;Note: the translated sequence in GenBank entry HUMPDHE1B, release 114.0, (PIDN:AAA600
on
R;Huh, T.L.; Chi, Y.T.; Casazza, J.P.; Veech, R.L.; Song, B.J.
submitted to the EMBL Data Library, April 1990
A;Description: Identical sequences for human brain and liver pyruvate dehydrogenase E1'
A;Reference number: S11715
A;Accession: S11715
A;Molecule type: mRNA
A;Residues: 1-390 <HU1>
```

A;Cross-references: EMBL:X52709; NID:g35378; PIDN:CAA36933.1; PID:g35379
A;Experimental source: brain
A;Accession: S11716
A;Molecule type: mRNA
A;Residues: 1-390 <HU2>
A;Cross-references: EMBL:X52710; NID:g35380; PIDN:CAA36934.1; PID:g35381
A;Experimental source: liver
R;Koike, K.; Ohta, S.; Urata, Y.; Kagawa, Y.; Koike, M.
Proc. Natl. Acad. Sci. U.S.A. 85, 41-45, 1988
A;Title: Cloning and sequencing of cDNAs encoding alpha and beta subunits of human pyruv
A;Reference number: A94198; MUID:88124815; PMID:3422424
A;Accession: A28398
A;Status: significant sequence differences
A;Molecule type: mRNA
A;Cross-references: GB:J03575; NID:g189737; PIDN:AAA60050.1; PID:g189738
A;Note: the sequence is extensively revised in reference JQ0770; the revised sequence is
R;Dahl, H.H.M.; Hunt, S.M.; Hutchison, W.M.; Brown, G.K.
J. Biol. Chem. 262, 7398-7403, 1987
A;Title: The human pyruvate dehydrogenase complex. Isolation of cDNA clones for the E1-a
A;Reference number: A29577; MUID:87222349; PMID:3034892
A;Accession: A29577
A;Status: significant sequence differences
A;Molecule type: mRNA
A;Cross-references: GB:M27166; NID:g488487; PIDN:AAA60051.1; PID:g387009
A;Note: the sequence reported in this reference is incorrect due to multiple frameshift
R;Ito, M.; Huq, A.H.; Naito, E.; Saijo, T.; Takeda, E.; Kuroda, Y.
J. Inherit. Metab. Dis. 15, 848-856, 1992
A;Title: Mutation of E1 alpha gene in a female patient with pyruvate dehydrogenase defic
A;Reference number: I56309; MUID:93188403; PMID:1338114
A;Accession: I56309
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 374-382,'SVDQV' <ITO>
A;Cross-references: GB:S56181; NID:g299257; PIDN:AAB25674.1; PID:g299258
R;De Meirleir, L.; Lissens, W.; Vamos, E.; Liebaers, I.
Hum. Genet. 88, 649-652, 1992
A;Title: Pyruvate dehydrogenase (PDH) deficiency caused by a 21-base pair insertion muta
A;Reference number: I54262; MUID:92201830; PMID:1551669
A;Accession: I54262
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 300-305,'DSYRTRE',306-309 <DEM2>
A;Cross-references: GB:S89908; NID:g247722; PIDN:AAB21863.1; PID:g247723
R;Hansen, L.L.; Horn, N.; Dahl, H.H.; Kruse, T.A.
Hum. Mol. Genet. 3, 1021-1022, 1994
A;Title: Pyruvate dehydrogenase deficiency caused by a 33 base pair duplication in the P
A;Reference number: I54356; MUID:95038723; PMID:7545958
A;Accession: I54356
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 300-307,'PPHSYRTREEI',308-311 <HAN>
A;Cross-references: GB:S75842; NID:g861533; PIDN:AAB32351.1; PID:g861534
C;Comment: The phosphorylation of several serine sites by [pyruvate dehydrogenase (lipoa
mplex. The enzyme is dephosphorylated by [pyruvate dehydrogenase (lipoamide)]-phosphatas
C;Genetics:
A;Gene: GDB:PDHA1
A;Cross-references: GDB:118895; OMIM:312170
A;Map position: Xp22.1-Xp22.1
A;Introns: 19/3; 23/3; 97/3; 140/1; 170/3; 201/3; 253/3; 272/3; 300/2; 336/3
C;Complex: the E1 component (pyruvate dehydrogenase) is a heterotetramer of two alpha an
yltransferase) monomers and 6 E3 component (lipoamide dehydrogenase) dimers
C;Function:
A;Description: catalyzes the reaction of pyruvate and lipoamide (on the E2 component) to
of pyruvate by NAD to form acetyl-CoA, carbon dioxide and NADH
A;Pathway: pyruvate metabolism
A;Note: thiamin pyrophosphate is a cofactor
C;Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-bin
C;Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; phosphoprotein;
F;1-29/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;30-390/Product: pyruvate dehydrogenase (lipoamide) alpha chain #status predicted <MAT>
F;185-234/Domain: thiamin pyrophosphate-binding domain homology <TPB>
F;232,293,300/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 76.0%; Score 38; DB 1; Length 390;
Best Local Similarity 80.0%; Pred. No. 8.9;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
| |||:||||
Db 54 TVLTREDGLK 63

RESULT 6
DERTP1
pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) alpha chain 1 precursor - rat
N;Alternate names: pyruvate dehydrogenase complex, E1 component alpha chain
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: S21553; I56503
R;Cullingford, T.E.; Phillips, I.I.; Clark, J.J.
submitted to the EMBL Data Library, June 1992
A;Reference number: S21553
A;Accession: S21553
A;Molecule type: mRNA
A;Residues: 1-390 <CULS>
A;Cross-references: UNIPROT:P26284; EMBL:Z12158; NID:g57656; PIDN:CAA78146.1; PID:g5765
A;Experimental source: liver
R;Cullingford, T.E.; Clark, J.B.; Phillips, I.R.
J. Neurochem. 62, 1682-1690, 1994
A;Title: The pyruvate dehydrogenase complex: cloning of the rat somatic E1 alpha subuni
brain.
A;Reference number: I56503; MUID:94209873; PMID:8158120
A;Accession: I56503
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-390 <CULP>
A;Cross-references: EMBL:Z12158; NID:g57656; PIDN:CAA78146.1; PID:g57657
A;Note: in Genbank entry RRPDHYE1A, release 113.0, the source is designated as Rattus r
C;Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-bi
C;Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; phosphoprotein
F;1-29/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;30-390/Product: pyruvate dehydrogenase (lipoamide) alpha chain #status predicted <MAT
F;185-234/Domain: thiamin pyrophosphate-binding domain homology <TPB>
F;232/Binding site: phosphate (Ser) (covalent) #status predicted
F;293/Binding site: phosphate (Ser) (covalent) #status predicted
F;300/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 76.0%; Score 38; DB 1; Length 390;
Best Local Similarity 80.0%; Pred. No. 8.9;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDGLK 10
| |||:||||
Db 54 TVLTREDGLK 63

RESULT 7
DERTPA
pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) alpha chain precursor - rat
N;Alternate names: pyruvate dehydrogenase complex, E1 component alpha chain
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: S15891
R;Matuda, S.; Nakano, K.; Ohta, S.; Saheki, T.; Kawanishi, Y.; Miyata, T.
Biochim. Biophys. Acta 1089, 1-7, 1991
A;Title: The alpha-ketoacid dehydrogenase complexes. Sequence similarity of rat pyruvat
A;Reference number: S15891; MUID:91223087; PMID:2025639
A;Accession: S15891
A;Molecule type: mRNA
A;Residues: 1-390 <MAT1>
A;Cross-references: UNIPROT:P26284
C;Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-bi
C;Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; phosphoprotein
F;1-29/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;30-390/Product: pyruvate dehydrogenase (lipoamide) alpha chain #status predicted <MAT
F;185-234/Domain: thiamin pyrophosphate-binding domain homology <TPB>

```
F;232/Binding site: phosphate (Ser) (covalent) #status predicted
F;293/Binding site: phosphate (Ser) (covalent) #status predicted
F;300/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             76.0%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   80.0%;  Pred. No. 8.9;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         54 TVLTREDGLK 63


RESULT 8
S23506
pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
C;Accession: S23506
R;Fitzgerald, J.; Hutchison, W.M.; Dahl, H.H.M.
Biochim. Biophys. Acta 1131, 83-90, 1992
A;Title: Isolation and characterisation of the mouse pyruvate dehydrogenase E1alpha gene
A;Reference number: S23506; MUID:92256495; PMID:1581363
A;Accession: S23506
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-390 <FIT>
A;Cross-references: UNIPROT:P35486; EMBL:M76727; NID:g200276; PIDN:AAA53046.1; PID:g2002
C;Genetics:
A;Gene: pdha-1
C;Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-bin
C;Keywords: mitochondrion; oxidoreductase; phosphoprotein
F;185-234/Domain: thiamin pyrophosphate-binding domain homology <TPB>

  Query Match             76.0%;  Score 38;  DB 2;  Length 390;
  Best Local Similarity   80.0%;  Pred. No. 8.9;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         54 TVLTREDGLK 63


RESULT 9
JH0102
apolipoprotein B - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: JH0102
R;Smith, T.J.
submitted to GenBank, June 1990
A;Reference number: A38864
A;Accession: JH0102
A;Molecule type: DNA
A;Residues: 1-779 <SMI>
A;Cross-references: UNIPROT:Q60536; GB:M35187
A;Note: this is a revision to the sequence from reference JH0101
R;Smith, T.J.; Hautamaa, D.; Maeda, N.
Gene 87, 309-310, 1990
A;Title: Sequence of the putative low-density lipoprotein receptor-binding regions of ap
A;Reference number: JH0101; MUID:90236327; PMID:2332175
A;Contents: annotation
A;Note: this sequence has been revised in reference A38864
C;Genetics:
A;Gene: apoB
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;
F;435-445/Region: receptor binding
F;646-656/Region: receptor binding

  Query Match             76.0%;  Score 38;  DB 2;  Length 779;
  Best Local Similarity   80.0%;  Pred. No. 17;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              :|||||  |||
Db        642 SRLTRKRGLK 651


RESULT 10
E60950
apolipoprotein B-100 - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: E60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LD
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: E60950
A;Molecule type: mRNA
A;Residues: 1-275 <LAW>
A;Cross-references: UNIPROT:Q7LZ77
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoproteir

  Query Match             72.0%;  Score 36;  DB 2;  Length 275;
  Best Local Similarity   80.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | ||||  |||
Db        221 TSLTRKRGLK 230


RESULT 11
T46858
molybdenum cofactor biosynthesis protein A [imported] - Rhodobacter sphaeroides
C;Species: Rhodobacter sphaeroides
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 12-Jul-2004
C;Accession: T46858
R;Mackenzie, C.; Simmons, A.E.; Kaplan, S.
Genetics 153, 525-538, 1999
A;Title: Multiple chromosomes in bacteria. The Yin and Yang of trp gene localization in
A;Reference number: Z24108; MUID:99442363; PMID:10511537
A;Accession: T46858
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-394 <MAC>
A;Cross-references: UNIPROT:Q9ZFA5; EMBL:AF108766; NID:g4185542; PIDN:AAD09121.1; PID:g
A;Experimental source: strain 2.4.1
C;Genetics:
A;Gene: moeA
A;Map position: I
C;Superfamily: molybdenum cofactor biosynthesis protein, MoeA type

  Query Match             72.0%;  Score 36;  DB 2;  Length 394;
  Best Local Similarity   87.5%;  Pred. No. 23;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDGL 9
              ||||:|||
Db        341 RLTRQDGL 348


RESULT 12
A84109
ABC transporter (ATP-binding protein) BH3673 [imported] - Bacillus halodurans (strain C
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: A84109
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: A84109
```

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-227 <STO>
A;Cross-references: UNIPROT:Q9K6Q5; GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAB073
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3673
C;Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homology

  Query Match             70.0%;  Score 35;  DB 2;  Length 227;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LTRKDGL 9
              |||||||
Db         70 LTRKDGL 76


RESULT 13
T29604
hypothetical protein ZK816.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T29604
R;Nhan, M.; Le, T.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid ZK816.
A;Reference number: Z20649
A;Accession: T29604
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-325 <NHA>
A;Cross-references: UNIPROT:Q23612; EMBL:U41018; PIDN:AAA82327.1; CESP:ZK816.5
C;Genetics:
A;Gene: CESP:ZK816.5
A;Introns: 24/1; 111/3; 170/2; 228/3; 280/1

  Query Match             70.0%;  Score 35;  DB 2;  Length 325;
  Best Local Similarity   60.0%;  Pred. No. 30;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              ||: ||:|:|
Db        148 TRIMRKNGMK 157


RESULT 14
G71050
asparagine synthase (glutamine-hydrolysing) (EC 6.3.5.4) [similarity] - Pyrococcus horik
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: G71050
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71050
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-480 <KAW>
A;Cross-references: UNIPROT:O58829; GB:AP000005; NID:g3236132; PIDN:BAA30201.1; PID:d103
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1102
C;Keywords: asparagine biosynthesis; ligase
F;2/Active site: Cys #status predicted

  Query Match             70.0%;  Score 35;  DB 2;  Length 480;
  Best Local Similarity   60.0%;  Pred. No. 44;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |:| |:||:|
Db        329 TKLAREDGVK 338


RESULT 15
AI1540
hypothetical protein lin0865 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AI1540
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1540
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-131 <GLA>
A;Cross-references: UNIPROT:Q92DE8; GB:AL592022; PIDN:CAC96097.1; PID:g16413316; GSPDB:
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0865

  Query Match             68.0%;  Score 34;  DB 2;  Length 131;
  Best Local Similarity   66.7%;  Pred. No. 20;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGL 9
              ||: ||||:
Db         74 TRIQRKDGV 82


Search completed: January 13, 2005, 01:52:36
Job time : 16.4262 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:15:45 ; Search time 78.0328 Seconds
(without alignments)
73.735 Million cell updates/sec

Title: US-09-823-418-10
Perfect score: 50
Sequence: 1 TRLTRKDGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	84.0	414	2	Q7YQR5	Q7yqr5 aotus vocif
2	42	84.0	596	2	Q28473	macaca fasc
3	42	84.0	675	2	Q90VV5	Q90vv5 fugu rubrip
4	42	84.0	678	2	P79997	P79997 oryzias lat
5	42	84.0	678	2	Q7T041	Q7t041 oryzias cur
6	42	84.0	678	2	Q9PWI2	Q9pwi2 oryzias lat
7	42	84.0	3262	2	Q13788	Q13788 homo sapien
8	42	84.0	4563	1	APB_HUMAN	P04114 homo sapien
9	42	84.0	4563	2	Q7Z600	Q7z600 homo sapien
10	38	76.0	89	2	Q9N1X8	Q9n1x8 bos taurus
11	38	76.0	389	1	ODPA_PIG	P29804 sus scrofa
12	38	76.0	390	1	ODPA_HUMAN	P08559 homo sapien
13	38	76.0	390	1	ODPA_MOUSE	P35486 mus musculu
14	38	76.0	390	1	ODPA_RAT	P26284 rattus norv
15	38	76.0	390	2	Q8HXW9	Q8hxw9 macaca fasc
16	38	76.0	390	2	AAH07142	Aah07142 mus muscu
17	38	76.0	393	2	Q6DGZ9	Q6dgz9 brachydanio
18	38	76.0	421	2	Q7TN68	Q7tn68 glaucomys v
19	38	76.0	432	2	Q7YR10	Q7yr10 diceros bic
20	38	76.0	436	2	Q7YQM8	Q7yqm8 nyctimene a
21	38	76.0	438	2	Q7YQM7	Q7yqm7 pteropus hy
22	38	76.0	438	2	Q7YR04	Q7yr04 rousettus a
23	38	76.0	445	2	Q7YR08	Q7yr08 chaetophrac
24	38	76.0	445	2	Q7TN64	Q7tn64 agouti paca
25	38	76.0	445	2	Q7TN71	Q7tn71 hydrochoeru
26	38	76.0	445	2	Q7TN72	Q7tn72 erethizon d
27	38	76.0	780	2	Q60536	Q60536 mesocricetu
28	38	76.0	780	2	Q60537	Q60537 mesocricetu
29	37	74.0	446	2	Q7ULC1	Q7ulc1 rhodopirell
30	37	74.0	6995	2	Q96RK2	Q96rk2 homo sapien
31	37	74.0	22152	2	Q8WXI7	Q8wxi7 homo sapien
32	36	72.0	201	2	O95211	O95211 homo sapien
33	36	72.0	202	1	SECU_HUMAN	O95997 homo sapien
34	36	72.0	202	2	CAG33416	Cag33416 homo sapi
35	36	72.0	208	2	Q8P9V7	Q8p9v7 xanthomonas
36	36	72.0	208	2	Q8PLN6	Q8pln6 xanthomonas
37	36	72.0	275	2	Q7LZ77	Q7lz77 gallus gall
38	36	72.0	387	2	Q7YQN2	Q7yqn2 phalanger o
39	36	72.0	394	2	Q9ZFA5	Q9zfa5 rhodobacter
40	36	72.0	400	2	Q7YQM9	Q7yqm9 ornithorhyn
41	36	72.0	405	2	Q7YQN0	Q7yqn0 tachyglossu
42	36	72.0	445	2	Q7TN70	Q7tn70 dinomys bra
43	36	72.0	452	2	Q73L73	Q73l73 treponema d
44	36	72.0	452	2	AAS12506	Aas12506 treponema
45	36	72.0	669	2	Q8AAW5	Q8aaw5 bacteroides

ALIGNMENTS

RESULT 1
Q7YQR5
ID Q7YQR5 PRELIMINARY; PRT; 414 AA.
AC Q7YQR5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Apolipoprotein B 100 (Fragment).
GN Name=apoB-100;
OS Aotus vociferans (Spix's owl monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX NCBI_TaxID=57176;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22761261; PubMed=12878460;
RA Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT "A new phylogenetic marker, apolipoprotein B, provides compelling
RT evidence for eutherian relationships.";
RL Mol. Phylogenet. Evol. 28:225-240(2003).
DR EMBL; AF548396; AAP97352.1; -.
KW Lipoprotein.
FT NON_TER 1 1
FT NON_TER 414 414
SQ SEQUENCE 414 AA; 45955 MW; EEFA8492157E1BDE CRC64;

Query Match 84.0%; Score 42; DB 2; Length 414;
Best Local Similarity 90.0%; Pred. No. 5.8;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TRLTRKDGLK 10
              |||||| |||
Db        258 TRLTRKRGLK 267
```

RESULT 2
Q28473
ID Q28473 PRELIMINARY; PRT; 596 AA.
AC Q28473;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Apolipoprotein B (Fragment).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92075708; PubMed=1742325;
RA Pape M.E., Castle C.K., Murray R.W., Funk G.M., Hunt C.E.,
RA Marotti K.R., Melchior G.W.;

```
RT   "Apo B metabolism in the cynomolgus monkey: evidence for post-
RT   transcriptional regulation.";
RL   Biochim. Biophys. Acta 1086:326-334(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Murray R.;
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X15737; CAA33755.1; -.
DR   PIR; S32802; S32802.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     596    596
SQ   SEQUENCE   596 AA;  66757 MW;  B13BBA74E25C3120 CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 596;
  Best Local Similarity   90.0%;  Pred. No. 8.6;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db        226 TRLTRKRGLK 235


RESULT 3
Q90VV5
ID   Q90VV5      PRELIMINARY;      PRT;   675 AA.
AC   Q90VV5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Soluble guanylyl cyclase alpha1 subunit.
GN   Name=FrGCS-alpha1;
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Morinaga C., Yamamoto T., Moriya Y., Suzuki N.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB062171; BAB60907.1; -.
DR   EMBL; AB062169; BAB60905.1; -.
DR   HSSP; P30803; 1AZS.
DR   GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF00211; Guanylate_cyc; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase.
SQ   SEQUENCE   675 AA;  75498 MW;  E71A283DC0369601 CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 675;
  Best Local Similarity   88.9%;  Pred. No. 9.9;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDGLK 10
              ||||||||:
Db        315 RLTRKDGLR 323


RESULT 4
P79997
ID   P79997      PRELIMINARY;      PRT;   678 AA.
AC   P79997;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Soluble guanylyl cyclase alpha subunit (EC 4.6.1.2).
OS   Oryzias latipes (Medaka fish) (Japanese ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98237571; PubMed=9578459;
RA   Mikami T., Kusakabe T., Suzuki N.;
RT   "Molecular cloning of cDNAs and expression of mRNAs encoding alpha and
RT   beta subunits of soluble guanylyl cyclase from medaka fish Oryzias
RT   latipes.";
RL   Eur. J. Biochem. 253:42-48(1998).
DR   EMBL; AB000849; BAA19198.1; -.
DR   HSSP; P30803; 1AZS.
DR   GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF00211; Guanylate_cyc; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase.
SQ   SEQUENCE   678 AA;  75133 MW;  3A6141D2A6475D40 CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 678;
  Best Local Similarity   88.9%;  Pred. No. 9.9;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDGLK 10
              ||||||||:
Db        316 RLTRKDGLR 324


RESULT 5
Q7T041
ID   Q7T041      PRELIMINARY;      PRT;   678 AA.
AC   Q7T041;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Soluble guanylyl cyclase alpha1 subunit.
GN   Name=OcGCS-alpha1;
OS   Oryzias curvinotus (Hynann ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=104658;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Shiga T., Suzuki N.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB115703; BAC80220.1; -.
DR   GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF00211; Guanylate_cyc; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase.
SQ   SEQUENCE   678 AA;  75128 MW;  C03C174B535ACCD9 CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 678;
  Best Local Similarity   88.9%;  Pred. No. 9.9;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 RLTRKDGLK 10
              |||||||||:
Db        316 RLTRKDGLR 324


RESULT 6
Q9PWI2
ID   Q9PWI2      PRELIMINARY;      PRT;   678 AA.
AC   Q9PWI2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Soluble guanylyl cyclase alpha subunit.
OS   Oryzias latipes (Medaka fish) (Japanese ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99303623; PubMed=10373466;
RA   Mikami T., Kusakabe T., Suzuki N.;
RT   "Tandem organization of medaka fish soluble guanylyl cylclase alpha1
RT   and beta1 subunit genes. Implications for coordinated transcription of
RT   two subunit genes.";
RL   J. Biol. Chem. 274:18567-18573(1999).
DR   EMBL; AB022280; BAA76690.1; -.
DR   HSSP; P30803; 1AZS.
DR   GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF00211; Guanylate_cyc; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase.
SQ   SEQUENCE   678 AA;  75165 MW;  66623709CBE68A5C CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 678;
  Best Local Similarity   88.9%;  Pred. No. 9.9;
  Matches    8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDGLK 10
              |||||||||:
Db        316 RLTRKDGLR 324


RESULT 7
Q13788
ID   Q13788      PRELIMINARY;      PRT;  3262 AA.
AC   Q13788;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   APOB protein (Fragment).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87191999; PubMed=2883086;
RA   Carlsson P., Darnfors C., Olofsson S.O., Bjursell G.;
RT   "Analysis of the human apolipoprotein B gene; complete structure of
RT   the B-74 region.";
RL   Gene 49:29-51(1986).
DR   EMBL; M15421; AAA51758.1; -.
DR   PIR; A27850; LPHUB.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0005319; F:lipid transporter activity; NAS.
DR   GO; GO:0006869; P:lipid transport; NAS.
FT   NON_TER       1      1
SQ   SEQUENCE   3262 AA;  370140 MW;  56603BC0618DD40D CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 3262;
  Best Local Similarity   90.0%;  Pred. No. 56;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db       2084 TRLTRKRGLK 2093


RESULT 8
APB_HUMAN
ID   APB_HUMAN      STANDARD;      PRT;  4563 AA.
AC   P04114; O00502; Q13787;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
DE   B-48 (Apo B-48)].
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87016385; PubMed=3763409;
RA   Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
RA   Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
RT   "Complete cDNA and derived protein sequence of human apolipoprotein B-
RT   100.";
RL   Nucleic Acids Res. 14:7501-7503(1986).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT GLU-4181.
RX   MEDLINE=88003974; PubMed=3652907;
RA   Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
RA   Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
RT   "DNA sequence of the human apolipoprotein B gene.";
RL   DNA 6:363-372(1987).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS ILE-98 AND GLU-4181.
RX   MEDLINE=87008488; PubMed=3759943;
RA   Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
RA   Gotto A.M. Jr., Chan L.;
RT   "The complete cDNA and amino acid sequence of human apolipoprotein B-
RT   100.";
RL   J. Biol. Chem. 261:12918-12921(1986).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87041416; PubMed=3464946;
RA   Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
RA   Lee N., Brewer H.B. Jr.;
RT   "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
RT   derived amino acid sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87161758; PubMed=3030729;
RA   Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
RA   Zannis V.I.;
RT   "The complete sequence and structural analysis of human apolipoprotein
RT   B-100: relationship between apoB-100 and apoB-48 forms.";
RL   EMBO J. 5:3495-3507(1986).
RN   [6]
RP   SEQUENCE OF 709-906 FROM N.A.
RX   MEDLINE=85270450; PubMed=3860836;
RA   Deeb S.S., Motulsky A.G., Albers J.J.;
RT   "A partial cDNA clone for human apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN   [7]
```

RP SEQUENCE OF 3056-3159 FROM N.A.
RX MEDLINE=86041888; PubMed=3903660;
RA Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
RA Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
RT "Human apolipoprotein B: identification of cDNA clones and
RT characterization of mRNA.";
RL Nucleic Acids Res. 13:6937-6953(1985).
RN [8]
RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A., AND VARIANT GLU-4181.
RX MEDLINE=86093680; PubMed=3841204;
RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
RA Bjursell G.;
RT "Molecular cloning of human apolipoprotein B cDNA.";
RL Nucleic Acids Res. 13:8813-8826(1985).
RN [9]
RP SEQUENCE OF 3109-4563 FROM N.A.
RX MEDLINE=85300528; PubMed=2994225;
RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Urdea M.S.,
RA Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M.,
RA Priestley L.M., Robertson E., Rall L.B., Betsholtz C., Shows T.B.,
RA Mahley R.W., Scott J.;
RT "Human apolipoprotein B: structure of carboxyl-terminal domains, sites
RT of gene expression, and chromosomal localization.";
RL Science 230:37-43(1985).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
RT "Isolation of a cDNA clone encoding the amino-terminal region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A., AND VARIANT ILE-98.
RX MEDLINE=86287319; PubMed=3461454;
RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanaka M.,
RA Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT "Analysis of cDNA clones encoding the entire B-26 region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX MEDLINE=88018019; PubMed=3659919;
RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
RA Gotto A.M. Jr., Li W.-H., Chan L.;
RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA Levy-Wilson B., Scott J.;
RT "Complete protein sequence and identification of structural domains of
RT human apolipoprotein B.";
RL Nature 323:734-738(1986).
RN [14]
RP DOMAINS.
RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT "Sequence, structure, receptor-binding domains and internal repeats of
RT human apolipoprotein B-100.";
RL Nature 323:738-742(1986).
RN [15]
RP CALCIUM-BINDING DATA.
RX MEDLINE=86242245; PubMed=3087360;
RA Dashti N., Lee D.M., Mok T.;
RT "Apolipoprotein B is a calcium binding protein.";
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN [16]
RP PALMITOYLATION OF CYS-1112.
RX MEDLINE=20143590; PubMed=10679026;
RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT "Palmitoylation of apolipoprotein B is required for proper
RT intracellular sorting and transport of cholesteroyl esters and
RT triglycerides.";
RL Mol. Biol. Cell 11:721-734(2000).
RN [17]
RP VARIANT SER-4338.
RX MEDLINE=91071750; PubMed=1979313;
RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA Cunny G., Cambien F., Roizes G.;
RT "Detection by denaturing gradient gel electrophoresis of a new
RT polymorphism in the apolipoprotein B gene.";
RL Hum. Genet. 86:91-93(1990).
RN [18]
RP VARIANT FDB GLN-3527.
RX MEDLINE=89098975; PubMed=2563166;
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA McCarthy B.J.;
RT "Association between a specific apolipoprotein B mutation and familial
RT defective apolipoprotein B-100.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN [19]
RP VARIANT LEU-2739.
RX MEDLINE=91016974; PubMed=2216805;
RA Huang L.-S., Gavish D., Breslow J.L.;
RT "Sequence polymorphism in the human apoB gene at position 8344.";
RL Nucleic Acids Res. 18:5922-5922(1990).
RN [20]
RP VARIANT FDB CYS-3558.
RX MEDLINE=95190020; PubMed=7883971;
RA Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT "Familial ligand-defective apolipoprotein B. Identification of a new
RT mutation that decreases LDL receptor binding affinity.";
RL J. Clin. Invest. 95:1225-1234(1995).
RN [21]
RP VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP AND THR-4481.
RX MEDLINE=97044521; PubMed=8889592;
RA Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT PCR-SSCP.";
RL Hum. Mutat. 8:282-285(1996).
RN [22]
RP VARIANTS FDB GLN-3527 AND CYS-3558.
RX MEDLINE=97403938; PubMed=9259199;
RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA Krempf M., Giraudet P., Junien C., Boileau C.;
RT "Familial ligand-defective apolipoprotein B-100: simultaneous
RT detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a French
RT population.";
RL Hum. Mutat. 10:160-163(1997).
RN [23]
RP VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP AND ILE-3921.
RX MEDLINE=98141125; PubMed=9490296;
RA Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT "Screening for mutations of the apolipoprotein B gene causing
RT hypocholesterolemia.";
RL Hum. Genet. 102:44-49(1998).
CC -!- FUNCTION: Apolipoprotein B is a major protein constituent of
CC chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo
CC B-100 functions as a recognition signal for the cellular binding
CC and internalization of LDL particles by the apoB/E receptor.
CC -!- SUBCELLULAR LOCATION: Secreted.

Query Match 84.0%; Score 42; DB 1; Length 4563;
Best Local Similarity 90.0%; Pred. No. 81;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TRLTRKDGLK 10
              |||||| |||
Db       3385 TRLTRKRGLK 3394


RESULT 9
Q7Z600
ID   Q7Z600      PRELIMINARY;      PRT;  4563 AA.
AC   Q7Z600;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Apolipoprotein B (Including Ag(X) antigen).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA   Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY324608; AAP72970.1; -.
DR   GO; GO:0005319; F:lipid transporter activity; IEA.
DR   GO; GO:0006869; P:lipid transport; IEA.
DR   InterPro; IPR009454; DUF1081.
DR   InterPro; IPR001747; Lipid_transprt_N.
DR   Pfam; PF06448; DUF1081; 1.
DR   Pfam; PF01347; Vitellogenin_N; 1.
DR   SMART; SM00638; LPD_N; 1.
KW   Lipoprotein.
SQ   SEQUENCE   4563 AA;  515553 MW;  030B34167CEDC63C CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 81;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              |||||| |||
Db       3385 TRLTRKRGLK 3394


RESULT 10
Q9N1X8
ID   Q9N1X8      PRELIMINARY;      PRT;    89 AA.
AC   Q9N1X8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Pyruvate dehydrogenase E1 component alpha subunit (Fragment).
GN   Name=PDHA;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Poloumienko A., Blecher S.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF132072; AAF36517.1; -.
DR   PIR; A60225; A60225.
DR   HSSP; P08559; 1NI4.
DR   GO; GO:0016624; F:oxidoreductase activity, acting on the alde. . .; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR001017; Dehydrogenase_E1.
DR   Pfam; PF00676; E1_dh; 1.
KW   Pyruvate.
FT   NON_TER       1      1
FT   NON_TER      89     89
SQ   SEQUENCE   89 AA;  10293 MW;  B47759C2D169292B CRC64;

  Query Match             76.0%;  Score 38;  DB 2;  Length 89;
  Best Local Similarity   80.0%;  Pred. No. 7.3;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         45 TVLTREDGLK 54


RESULT 11
ODPA_PIG
ID   ODPA_PIG       STANDARD;      PRT;   389 AA.
AC   P29804;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Pyruvate dehydrogenase E1 component alpha subunit, somatic form,
DE   mitochondrial precursor (EC 1.2.4.1) (PDHE1-A type I) (Fragment).
GN   Name=PDHA1;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=90370488; PubMed=2395657;
RA   Sermon K., Demeirleir L., Elpers I., Lissens W., Liebaers I.;
RT   "Characterisation of a cDNA for porcine PDH-E1 alpha and comparison
RT   with the human cDNA.";
RL   Nucleic Acids Res. 18:4925-4925(1990).
CC   -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
CC       conversion of pyruvate to acetyl-CoA and CO(2). It contains
CC       multiple copies of three enzymatic components: pyruvate
CC       dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
CC       lipoamide dehydrogenase (E3).
CC   -!- CATALYTIC ACTIVITY: Pyruvate + [dihydrolipoyllysine-residue
CC       acetyltransferase] lipoyllysine = [dihydrolipoyllysine-residue
CC       acetyltransferase] S-acetyldihydrolipoyllysine + CO(2).
CC   -!- COFACTOR: Thiamine pyrophosphate.
CC   -!- ENZYME REGULATION: E1 activity is regulated by phosphorylation
CC       (inactivation) and dephosphorylation (activation) of the alpha
CC       subunit.
CC   -!- SUBUNIT: Tetramer of two alpha and two beta subunits.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52990; CAA37180.1; -.
DR   PIR; S20813; DEPGPA.
DR   HSSP; P08559; 1NI4.
DR   InterPro; IPR001017; Dehydrogenase_E1.
DR   Pfam; PF00676; E1_dh; 1.
KW   Flavoprotein; Glycolysis; Mitochondrion; Oxidoreductase;
KW   Phosphorylation; Thiamine pyrophosphate; Transit peptide.
FT   NON_TER       1      1
FT   TRANSIT      <1     28       Mitochondrion (By similarity).
FT   CHAIN        29    389       Pyruvate dehydrogenase E1 component alpha
FT                                subunit, somatic form.
FT   MOD_RES     231    231       Phosphoserine (By similarity).
FT   MOD_RES     292    292       Phosphoserine (By similarity).
FT   MOD_RES     299    299       Phosphoserine (By similarity).
SQ   SEQUENCE   389 AA;  43121 MW;  E9C7DF85389A9A47 CRC64;
```

```
  Query Match           76.0%;  Score 38;  DB 1;  Length 389;
  Best Local Similarity   80.0%;  Pred. No. 37;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         53 TVLTREDGLK 62


RESULT 12
ODPA_HUMAN
ID   ODPA_HUMAN     STANDARD;      PRT;   390 AA.
AC   P08559; Q9NP12;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-OCT-2004 (Rel. 45, Last annotation update)
DE   Pyruvate dehydrogenase E1 component alpha subunit, somatic form,
DE   mitochondrial precursor (EC 1.2.4.1) (PDHE1-A type I).
GN   Name=PDHA1; Synonyms=PHE1A;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leukocyte;
RX   MEDLINE=91033044; PubMed=2227443;
RA   Koike K., Urata Y., Matsuo S., Koike M.;
RT   "Characterization and nucleotide sequence of the gene encoding the
RT   human pyruvate dehydrogenase alpha-subunit.";
RL   Gene 93:307-311(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89315791; PubMed=2748588;
RA   Ho L., Wexler I.D., Liu T.C., Thekkumkara T.J., Patel M.S.;
RT   "Characterization of cDNAs encoding human pyruvate dehydrogenase alpha
RT   subunit.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:5330-5334(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Liver;
RA   Huh T.L., Chi Y.T., Casazza J.P., Veech R.L., Song B.J.;
RL   Submitted (APR-1990) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87222349; PubMed=3034892;
RA   Dahl H.-H.M., Hunt S.M., Hutchison W.M., Brown G.K.;
RT   "The human pyruvate dehydrogenase complex. Isolation of cDNA clones
RT   for the E1 alpha subunit, sequence analysis, and characterization of
RT   the mRNA.";
RL   J. Biol. Chem. 262:7398-7403(1987).
RN   [5]
RP   REVISIONS.
RX   MEDLINE=89308653; PubMed=2745444;
RA   Maragos C., Hutchinson W.M., Hayasaki K., Brown G.K., Dahl H.-H.M.;
RT   "Structural organization of the gene for the E1 alpha subunit of the
RT   human pyruvate dehydrogenase complex.";
RL   J. Biol. Chem. 264:12294-12298(1989).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88115327; PubMed=2828359;
RA   de Meirleir L., MacKay N., Wah A.M.L.H., Robinson B.H.;
RT   "Isolation of a full-length complementary DNA coding for human E1
RT   alpha subunit of the pyruvate dehydrogenase complex.";
RL   J. Biol. Chem. 263:1991-1995(1988).
RN   [7]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88124815; PubMed=3422424;
RA   Koike K., Ohta S., Urata Y., Kagawa Y., Koike M.;
RT   "Cloning and sequencing of cDNAs encoding alpha and beta subunits of
RT   human pyruvate dehydrogenase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:41-45(1988).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [9]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=93244853; PubMed=1301207;
RA   Dahl H.-H.M., Brown G.K., Brown R.M., Hansen L.L., Kerr D.S.,
RA   Wexler I.D., Patel M.S., de Meirleir L., Lissens W., Chun K.,
RA   McKay N., Robinson B.H.;
RT   "Mutations and polymorphisms in the pyruvate dehydrogenase E1 alpha
RT   gene.";
RL   Hum. Mutat. 1:97-102(1992).
RN   [10]
RP   VARIANT PDHE1 DEFICIENCY LYS-313 DEL, AND VARIANT LS HIS-378.
RX   MEDLINE=91359689; PubMed=1909401;
RA   Hansen L.L., Brown G.K., Kirby D.M., Dahl H.-H.M.;
RT   "Characterization of the mutations in three patients with pyruvate
RT   dehydrogenase E1 alpha deficiency.";
RL   J. Inherit. Metab. Dis. 14:140-151(1991).
RN   [11]
RP   VARIANT PDHE1 DEFICIENCY CYS-302.
RX   MEDLINE=93188402; PubMed=1293379;
RA   Dahl H.-H.M., Hansen L.L., Brown R.M., Danks D.M., Rogers J.G.,
RA   Brown G.K.;
RT   "X-linked pyruvate dehydrogenase E1 alpha subunit deficiency in
RT   heterozygous females: variable manifestation of the same mutation.";
RL   J. Inherit. Metab. Dis. 15:835-847(1992).
RN   [12]
RP   VARIANT LS ALA-258.
RX   MEDLINE=93270474; PubMed=8498846;
RA   Matthews P.M., Marchington D.R., Squier M., Land J., Brown R.M.,
RA   Brown G.K.;
RT   "Molecular genetic characterization of an X-linked form of Leigh's
RT   syndrome.";
RL   Ann. Neurol. 33:652-655(1993).
RN   [13]
RP   VARIANTS PDHE1 DEFICIENCY MET-167; THR-199; ALA-231; GLY-263 AND
RP   LEU-292.
RX   MEDLINE=93278396; PubMed=8504306;
RA   Chun K., McKay N., Petrova-Benedict R., Robinson B.H.;
RT   "Mutations in the X-linked E1 alpha subunit of pyruvate dehydrogenase
RT   leading to deficiency of the pyruvate dehydrogenase complex.";
RL   Hum. Mol. Genet. 2:449-454(1993).
RN   [14]
RP   VARIANT LS LEU-205.
RX   MEDLINE=94258164; PubMed=8199595;
RA   Dahl H.-H.M., Brown G.K.;
RT   "Pyruvate dehydrogenase deficiency in a male caused by a point
RT   mutation (F205L) in the E1 alpha subunit.";
RL   Hum. Mutat. 3:152-155(1994).
RN   [15]
RP   VARIANT PDHE1 DEFICIENCY GLN-263.
RX   MEDLINE=95056975; PubMed=7967473;
```

```
RA   Awata H., Endo F., Tanoue A., Kitano A., Matsuda I.;
RT   "Characterization of a point mutation in the pyruvate dehydrogenase E1
RT   alpha gene from two boys with primary lactic acidaemia.";
RL   J. Inherit. Metab. Dis. 17:189-195(1994).
RN   [16]
RP   VARIANTS PDHE1 DEFICIENCY CYS-72; GLY-263 AND ARG-311 DEL, AND
RP   VARIANTS LS LEU-205 AND HIS-378.
RX   MEDLINE=95193781; PubMed=7887409;
RA   Chun K., MacKay N., Petrova-Benedict R., Federico A., Fois A.,
RA   Cole D.E.C., Robertson E., Robinson B.H.;
RT   "Mutations in the X-linked E1 alpha subunit of pyruvate dehydrogenase:
RT   exon skipping, insertion of duplicate sequence, and missense mutations
RT   leading to the deficiency of the pyruvate dehydrogenase complex.";
RL   Am. J. Hum. Genet. 56:558-569(1995).
RN   [17]
RP   VARIANT PDHE1 DEFICIENCY PRO-10.
RX   MEDLINE=96029268; PubMed=7573035;
RA   Takakubo F., Cartwright P., Hoogenraad N., Thorburn D.R., Collins F.,
RA   Lithgow T., Dahl H.H.;
RT   "An amino acid substitution in the pyruvate dehydrogenase E1 alpha
RT   gene, affecting mitochondrial import of the precursor protein.";
RL   Am. J. Hum. Genet. 57:772-780(1995).
RN   [18]
RP   VARIANT PDHE1 DEFICIENCY LEU-217.
RX   MEDLINE=95276751; PubMed=7757088;
RA   Hemalatha S.G., Kerr D.S., Wexler I.D., Lusk M.M., Kaung M., Du Y.,
RA   Kolli M., Schelper R.L., Patel M.S.;
RT   "Pyruvate dehydrogenase complex deficiency due to a point mutation
RT   (P188L) within the thiamine pyrophosphate binding loop of the E1 alpha
RT   subunit.";
RL   Hum. Mol. Genet. 4:315-318(1995).
RN   [19]
RP   VARIANTS PDHE1 DEFICIENCY CYS-72; ASP-113; ARG-162; GLY-263 AND
RP   CYS-302.
RX   MEDLINE=96263737; PubMed=8664900;
RA   Lissens W., de Meirleir L., Seneca S., Benelli C., Marsac C.,
RA   Poll-The B.T., Briones P., Ruitenbeek W., van Diggelen O., Chaigne D.,
RA   Ramaekers V., Liebaers I.;
RT   "Mutation analysis of the pyruvate dehydrogenase E1 alpha gene in
RT   eight patients with a pyruvate dehydrogenase complex deficiency.";
RL   Hum. Mutat. 7:46-51(1996).
RN   [20]
RP   VARIANTS PDHE1 DEFICIENCY VAL-210 AND ARG-311 DEL.
RX   MEDLINE=97001225; PubMed=8844217;
RA   Tripatara A., Kerr D.S., Lusk M.M., Kolli M., Tan J., Patel M.S.;
RT   "Three new mutations of the pyruvate dehydrogenase alpha subunit: a
RT   point mutation (M181V), 3 bp deletion (-R282), and 16 bp
RT   insertion/frameshift (K358SVS-->TVDQS).";
RL   Hum. Mutat. 8:180-182(1996).
RN   [21]
RP   VARIANTS PDHE1 DEFICIENCY CYS-302 AND HIS-302.
RX   MEDLINE=98334347; PubMed=9671272;
RA   Otero L.J., Brown R.M., Brown G.K.;
RT   "Arginine 302 mutations in the pyruvate dehydrogenase E1alpha subunit
RT   gene: identification of further patients and in vitro demonstration of
RT   pathogenicity.";
RL   Hum. Mutat. 12:114-121(1998).
CC   -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
CC       conversion of pyruvate to acetyl-CoA and CO(2). It contains
CC       multiple copies of three enzymatic components: pyruvate
CC       dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
CC       lipoamide dehydrogenase (E3).
CC   -!- CATALYTIC ACTIVITY: Pyruvate + [dihydrolipoyllysine-residue
CC       acetyltransferase] lipoyllysine = [dihydrolipoyllysine-residue
CC       acetyltransferase] S-acetyldihydrolipoyllysine + CO(2).

  Query Match             76.0%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   80.0%;  Pred. No. 37;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         54 TVLTREDGLK 63
```

```
RESULT 13
ODPA_MOUSE
ID   ODPA_MOUSE     STANDARD;      PRT;   390 AA.
AC   P35486;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Pyruvate dehydrogenase E1 component alpha subunit, somatic form,
DE   mitochondrial precursor (EC 1.2.4.1) (PDHE1-A type I).
GN   Name=Pdha1; Synonyms=Pdha-1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92256495; PubMed=1581363;
RA   Fitzgerald G.F., Hutchison W.M., Dahl H.H.M.;
RT   "Isolation and characterisation of the mouse pyruvate dehydrogenase E1
RT   alpha genes.";
RL   Biochim. Biophys. Acta 1131:83-90(1992).
CC   -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
CC       conversion of pyruvate to acetyl-CoA and CO(2). It contains
CC       multiple copies of three enzymatic components: pyruvate
CC       dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
CC       lipoamide dehydrogenase (E3).
CC   -!- CATALYTIC ACTIVITY: Pyruvate + [dihydrolipoyllysine-residue
CC       acetyltransferase] lipoyllysine = [dihydrolipoyllysine-residue
CC       acetyltransferase] S-acetyldihydrolipoyllysine + CO(2).
CC   -!- COFACTOR: Thiamine pyrophosphate.
CC   -!- ENZYME REGULATION: E1 activity is regulated by phosphorylation
CC       (inactivation) and dephosphorylation (activation) of the alpha
CC       subunit.
CC   -!- SUBUNIT: Tetramer of two alpha and two beta subunits.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC   -!- TISSUE SPECIFICITY: In all tissues, but in very low amount in
CC       testis.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76727; AAA53046.1; -.
DR   PIR; S23506; S23506.
DR   HSSP; P08559; 1NI4.
DR   MGD; MGI:97532; Pdha1.
DR   GO; GO:0004738; F:pyruvate dehydrogenase activity; IMP.
DR   InterPro; IPR001017; Dehydrogenase_E1.
DR   Pfam; PF00676; E1_dh; 1.
KW   Flavoprotein; Glycolysis; Mitochondrion; Multigene family;
KW   Oxidoreductase; Phosphorylation; Thiamine pyrophosphate;
KW   Transit peptide.
FT   TRANSIT      1     29       Mitochondrion (By similarity).
FT   CHAIN       30    390       Pyruvate dehydrogenase E1 component alpha
FT                               subunit, somatic form.
FT   MOD_RES     232    232       Phosphoserine (By similarity).
FT   MOD_RES     293    293       Phosphoserine (By similarity).
FT   MOD_RES     300    300       Phosphoserine (By similarity).
SQ   SEQUENCE   390 AA;  43231 MW;  40898944CE8E0A03 CRC64;

  Query Match             76.0%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   80.0%;  Pred. No. 37;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         54 TVLTREDGLK 63
```

```
RESULT 14
ODPA_RAT
ID   ODPA_RAT       STANDARD;      PRT;   390 AA.
AC   P26284;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Pyruvate dehydrogenase E1 component alpha subunit, somatic form,
DE   mitochondrial precursor (EC 1.2.4.1) (PDHE1-A type I).
GN   Name=Pdha1;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91223087; PubMed=2025639;
RA   Matuda S., Nakano K., Ohta S., Saheki T., Kawanishi Y., Miyata T.;
RT   "The alpha-ketoacid dehydrogenase complexes. Sequence similarity of
RT   rat pyruvate dehydrogenase with Escherichia coli and Azotobacter
RT   vinelandii alpha-ketoglutarate dehydrogenase.";
RL   Biochim. Biophys. Acta 1089:1-7(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=94209873; PubMed=8158120;
RA   Cullingford T.E., Clark J.B., Phillips I.R.;
RT   "The pyruvate dehydrogenase complex: cloning of the rat somatic E1
RT   alpha subunit and its coordinate expression with the mRNAs for the E1
RT   beta, E2, and E3 catalytic subunits in developing rat brain.";
RL   J. Neurochem. 62:1682-1690(1994).
CC   -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
CC       conversion of pyruvate to acetyl-CoA and CO(2). It contains
CC       multiple copies of three enzymatic components: pyruvate
CC       dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
CC       lipoamide dehydrogenase (E3).
CC   -!- CATALYTIC ACTIVITY: Pyruvate + [dihydrolipoyllysine-residue
CC       acetyltransferase] lipoyllysine = [dihydrolipoyllysine-residue
CC       acetyltransferase] S-acetyldihydrolipoyllysine + CO(2).
CC   -!- COFACTOR: Thiamine pyrophosphate.
CC   -!- ENZYME REGULATION: E1 activity is regulated by phosphorylation
CC       (inactivation) and dephosphorylation (activation) of the alpha
CC       subunit.
CC   -!- SUBUNIT: Tetramer of two alpha and two beta subunits.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC   -!- TISSUE SPECIFICITY: In all tissues, but in very low amount in
CC       testis.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z12158; CAA78146.1; -.
DR   PIR; S15891; DERTPA.
DR   PIR; S21553; DERTP1.
DR   HSSP; P08559; 1NI4.
DR   RGD; 3286; Pdha1.
DR   InterPro; IPR001017; Dehydrogenase_E1.
DR   Pfam; PF00676; E1_dh; 1.
KW   Flavoprotein; Glycolysis; Mitochondrion; Multigene family;
KW   Oxidoreductase; Phosphorylation; Thiamine pyrophosphate;
KW   Transit peptide.
FT   TRANSIT       1     29       Mitochondrion (By similarity).
FT   CHAIN        30    390       Pyruvate dehydrogenase E1 component alpha
FT                                 subunit, somatic form.
FT   MOD_RES     232    232       Phosphoserine (By similarity).
FT   MOD_RES     293    293       Phosphoserine (By similarity).
FT   MOD_RES     300    300       Phosphoserine (By similarity).
FT   CONFLICT     10     10       R -> H (in Ref. 2).
FT   CONFLICT    126    126       N -> T (in Ref. 2).
FT   CONFLICT    129    130       HA -> LP (in Ref. 2).
FT   CONFLICT    134    134       V -> I (in Ref. 2).
SQ   SEQUENCE   390 AA;  43212 MW;  21B78A8014460DC0 CRC64;

  Query Match             76.0%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   80.0%;  Pred. No. 37;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         54 TVLTREDGLK 63


RESULT 15
Q8HXW9
ID   Q8HXW9      PRELIMINARY;      PRT;   390 AA.
AC   Q8HXW9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Pyruvate dehydrogenase E1alpha.
GN   Name=pdha1;
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain cerebellum cortex;
RA   Kusuda J., Osada N., Hida M., Sugano S., Hashimoto K.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB083322; BAC20601.1; -.
DR   HSSP; P08559; 1NI4.
DR   GO; GO:0016624; F:oxidoreductase activity, acting on the alde. . .; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR001017; Dehydrogenase_E1.
DR   Pfam; PF00676; E1_dh; 1.
KW   Pyruvate.
SQ   SEQUENCE   390 AA;  43365 MW;  420CE38364BDB33C CRC64;

  Query Match             76.0%;  Score 38;  DB 2;  Length 390;
  Best Local Similarity   80.0%;  Pred. No. 37;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDGLK 10
              | |||:||||
Db         54 TVLTREDGLK 63


Search completed: January 13, 2005, 01:51:02
Job time : 80.0328 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:14:40 ; Search time 77.2131 Seconds
(without alignments)
46.460 Million cell updates/sec

Title: US-09-823-418-11
Perfect score: 49
Sequence: 1 TELTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	49	100.0	10	2	AAY30692	Aay30692 Apo-B100
2	46	93.9	10	2	AAY30693	Aay30693 Apo-B100
3	44	89.8	11	2	AAW57205	Aaw57205 Apo B bin
4	44	89.8	13	2	AAW57207	Aaw57207 Apo B 100
5	44	89.8	15	2	AAW41261	Aaw41261 Apolipopr
6	44	89.8	15	2	AAW96892	Aaw96892 ApoB-100
7	44	89.8	20	6	ABJ37575	Abj37575 Heparin b
8	44	89.8	22	2	AAW57208	Aaw57208 Apo B 100
9	44	89.8	22	2	AAW57209	Aaw57209 Apo B 100
10	44	89.8	34	5	AAE14541	Aae14541 Human apo
11	44	89.8	36	2	AAW96876	Aaw96876 Nucleic a
12	44	89.8	37	2	AAW64587	Aaw64587 Human apo
13	44	89.8	51	2	AAW96845	Aaw96845 Nucleic a
14	44	89.8	343	4	ABB37687	Abb37687 Peptide #
15	44	89.8	343	4	ABG52504	Abg52504 Human liv
16	44	89.8	377	2	AAR72704	Aar72704 Human apo
17	44	89.8	377	2	AAR34031	Aar34031 Sequence
18	44	89.8	2463	8	ADJ57400	Adj57400 Human apo
19	44	89.8	3923	2	AAY31237	Aay31237 Human Apo
20	44	89.8	4536	2	AAW41262	Aaw41262 Apolipopr
21	44	89.8	4536	2	AAW96826	Aaw96826 Amino aci
22	44	89.8	4560	5	AAU98981	Aau98981 Human apo
23	44	89.8	4561	7	ADD48677	Add48677 Human Pro
24	44	89.8	4563	5	AAO15893	Aao15893 Human apo
25	44	89.8	4563	6	ABR40253	Abr40253 Human ali
26	44	89.8	4563	6	ABU79140	Abu79140 Apolipopr
27	44	89.8	4563	7	ADF43408	Adf43408 Apolipopr
28	44	89.8	4563	8	ADH18871	Adh18871 Human apo
29	44	89.8	4563	8	ADH18870	Adh18870 Human apo
30	44	89.8	4563	8	ADO33445	Ado33445 Human apo
31	44	89.8	4563	8	ADO33447	Ado33447 Human apo
32	44	89.8	4590	4	AAU33184	Aau33184 Novel hum
33	41	83.7	465	4	AAB92994	Aab92994 Human pro
34	41	83.7	709	4	AAB95320	Aab95320 Human pro
35	41	83.7	719	7	ADM36086	Adm36086 Human acy
36	41	83.7	720	3	AAY94296	Aay94296 Human coe
37	41	83.7	720	5	ABG61883	Abg61883 Prostate
38	41	83.7	720	6	ABR63871	Abr63871 Human Fac
39	41	83.7	720	7	ADE31735	Ade31735 Human 266
40	41	83.7	720	7	ADN95365	Adn95365 Human BEC
41	41	83.7	720	8	ADJ45487	Adj45487 LXR-ligan
42	41	83.7	720	8	ADP05461	Adp05461 Human ace
43	40	81.6	10	2	AAY30682	Aay30682 Apo-B100
44	40	81.6	10	2	AAY30687	Aay30687 Apo-B100
45	39	79.6	10	2	AAY30690	Aay30690 Apo-B100

ALIGNMENTS

RESULT 1
AAY30692
ID AAY30692 standard; peptide; 10 AA.
XX
AC AAY30692;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can

```
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 10 AA;

  Query Match             100.0%;  Score 49;  DB 2;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.0095;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||||||||||
Db          1 TELTRKRGLK 10


RESULT 2
AAY30693
ID   AAY30693 standard; peptide; 10 AA.
XX
AC   AAY30693;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW   Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW   low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946598-A1.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US004805.
XX
PR   10-MAR-1998;   98US-0077618P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Innerarity TL,  Boren JOS;
XX
DR   WPI; 1999-551509/46.
XX
PT   Identifying compounds which affect binding of low density lipoprotein
PT   with proteoglycan, used for, e.g. obtaining compounds for reducing
PT   atherosclerosis.
XX
PS   Claim 17; Page 57; 70pp; English.
XX
CC   AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC   receptor mutations. They were created to identify compounds which
CC   modulate atherosclerosis. The peptides are derived from amino acids 3358
CC   to 3367 of apoB100. The method comprises detecting compounds which affect
CC   low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC   can be used for identifying compounds which disrupt LDL-PG binding
CC   without inhibiting LDL receptor binding. Such compounds can be used to
CC   reduce or prevent the formation of atherosclerotic lesions and prevent
CC   atherosclerosis. The transgenic non-human animals and mammals which
CC   express human apo-B100 can be used as an in vivo model system for the
CC   study of atherosclerosis, and in vivo assay methods for identifying
CC   compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 10 AA;

  Query Match             93.9%;  Score 46;  DB 2;  Length 10;
  Best Local Similarity   90.0%;  Pred. No. 0.037;
  Matches    9;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              |:||||||||
Db          1 TDLTRKRGLK 10


RESULT 3
AAW57205
ID   AAW57205 standard; peptide; 11 AA.
XX
AC   AAW57205;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B binding site peptide 2.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 12; Page 52; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B binding site
CC   peptide which can be used as a component of a non-naturally occurring,
CC   receptor-competent low density lipoprotein (LDL) particle of the present
CC   invention. The LDL particle comprises at least 1 peptide component that
CC   has at least 1 binding site for an apo B protein receptor and at least 1
CC   lipophilic substituent. Also described in the invention are peptides
CC   containing an apo B binding sequence with at least 70% identity with
CC   sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their dimers. Non-
CC   naturally occurring, receptor-competent LDL particles are useful as: (i)
CC   drug-targeting vectors for delivering anticancer drugs to cancer cells
CC   that express an apo B protein receptor, and (ii) additives for cell
CC   culture media especially as growth supplements. Non-naturally occurring,
CC   receptor-competent LDL particles do not require the complete apo B
CC   sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 11 AA;

  Query Match             89.8%;  Score 44;  DB 2;  Length 11;
  Best Local Similarity   90.0%;  Pred. No. 0.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db          2 TRLTRKRGLK 11
```

```
RESULT 4
AAW57207
ID   AAW57207 standard; peptide; 13 AA.
XX
AC   AAW57207;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide B.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 13 AA;

  Query Match             89.8%;  Score 44;  DB 2;  Length 13;
  Best Local Similarity   90.0%;  Pred. No. 0.12;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db          3 TRLTRKRGLK 12


RESULT 5
AAW41261
ID   AAW41261 standard; peptide; 15 AA.
XX
AC   AAW41261;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Apolipoprotein B-100 fragment.
XX
KW   Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW   thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
KW   angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW   prothrombinase complex.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9743311-A1.
XX
PD   20-NOV-1997.
XX
PF   09-MAY-1997;   97WO-GB001255.
XX
PR   09-MAY-1996;   96GB-00009702.
XX
PA   (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI   Bruckdorfer KR,  Ettelaie C;
XX
DR   WPI; 1998-008798/01.
XX
PT   Peptide fragments of apo:lipo:protein B-100 with anticoagulant activity -
PT   used for treating or preventing coagulation, inhibiting angiogenesis,
PT   cell differentiation and apoptosis.
XX
PS   Disclosure; Page 22; 60pp; English.
XX
CC   This sequence is an example of the peptide of the invention. It has the
CC   formula (I), or their variants with one or more internal deletions,
CC   insertions or substitutions, while retaining anti-coagulant properties of
CC   apolipoprotein B-100 (apoB-100). Z1-KAQ-X1-KKNKHRHS-X2-T-Z2 (I) X1 = S or
CC   Y; X2 = T or I; Z1 = the N terminus of the peptide, or 1-47 amino acids
CC   (aa); Z2 = the C terminus of the peptide, a terminal amide group or 1-77
CC   aa. Compositions containing the peptide are used for simultaneous,
CC   separate or sequential treatment of cancer, particularly to prevent
CC   metastatic spread. They are also used to inhibit thromboplastin-mediated
CC   processes, specifically to prevent or reduce blood coagulation (e.g.
CC   during or after surgery or in cases of heart attack, stroke etc.) and to
CC   inhibit angiogenesis, cellular differentiation or apoptosis. KRAD-14,
CC   which is active as such or as part of a 98-aa peptide, inhibits
CC   activation of the prothrombinase complex; and prevents activation of
CC   factor VII on the surface of thromboplastin and of platelets by thrombin.
CC   It binds to the residues 58-66 of thromboplastin. Since (I) are much
CC   smaller than apoB-100, they act more quickly
XX
SQ   Sequence 15 AA;

  Query Match             89.8%;  Score 44;  DB 2;  Length 15;
  Best Local Similarity   90.0%;  Pred. No. 0.14;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db          1 TRLTRKRGLK 10


RESULT 6
AAW96892
ID   AAW96892 standard; peptide; 15 AA.
XX
AC   AAW96892;
XX
DT   22-APR-1999  (first entry)
XX
DE   ApoB-100 nuclear localisation signal sequence, residues 3353-3367.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
```

KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Claim 19; Fig 13D; 293pp; English.
XX
CC AAW96878-97 represent nuclear localisation signal sequence derived from
CC human apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein
CC component of very-low density lipoproteins (VLDL), intermediate density
CC lipoprotein (IDL), low density lipoproteins ((LDL) and lipoprotein a. The
CC present sequence can be used in the composition of the invention. The
CC specification describes a composition that comprises LDL and
CC apolipoproteins for the binding and in vivo transport of nucleic acids.
CC The composition is used to deliver nucleic acids to eukaryotic cells, in
CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC molecule (or ribozyme). Specifically they are used for gene therapy of
CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC fibrosis and arteriosclerosis
XX
SQ Sequence 15 AA;

Query Match 89.8%; Score 44; DB 2; Length 15;
Best Local Similarity 90.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
| ||||||||
Db 6 TRLTRKRGLK 15

RESULT 7
ABJ37575
ID ABJ37575 standard; peptide; 20 AA.
XX
AC ABJ37575;
XX
DT 10-MAY-2003 (first entry)
XX
DE Heparin binding peptide sequence #28.
XX
KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW cardiovascular; circulatory; ligand; sulphated; sulphonated; tumour;
KW rheumatoid arthritis; hypoxia; diabetic retinopathy; heparin.
XX
OS Unidentified.
XX
PN WO2003007689-A2.
XX
PD 30-JAN-2003.
XX
PF 22-JUL-2002; 2002WO-US023419.
XX
PR 20-JUL-2001; 2001US-0306726P.
XX
PA (ETHZ-) ETH ZUERICH.
PA (UYZU-) UNIV ZURICH.
XX
PI Hubbell JA, Schoenmakers R, Maynard HD;
XX
DR WPI; 2003-300420/29.
XX
PT Use of a ligand comprising of at least one sulfated or sulfonated amino
PT acid for the treatment of e.g. tumors, rheumatoid arthritis, diabetic
PT retinopathy and hypoxia.
XX
PS Disclosure; Fig 2; 79pp; English.
XX
CC The invention relates to a novel ligand for binding a target biomolecule,
CC which comprises a peptide having at least one sulphated or sulphonated
CC amino acid and at least one amino acid chosen from neutral and positively
CC charged amino acids. The novel ligands can be used for the treatment of
CC e.g. tumours, rheumatoid arthritis, diabetic retinopathy and hypoxia.
CC This sequence represents a heparin binding peptide relating to the
CC invention
XX
SQ Sequence 20 AA;

Query Match 89.8%; Score 44; DB 6; Length 20;
Best Local Similarity 90.0%; Pred. No. 0.18;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
| ||||||||
Db 7 TRLTRKRGLK 16

RESULT 8
AAW57208
ID AAW57208 standard; peptide; 22 AA.
XX
AC AAW57208;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide C.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
FT Modified-site 22
FT /note= "attached to cholesterol"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.

XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 89.8%; Score 44; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
| ||||||||
Db 7 TRLTRKRGLK 16

RESULT 9
AAW57209
ID AAW57209 standard; peptide; 22 AA.
XX
AC AAW57209;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide D.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 89.8%; Score 44; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 0.2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
| ||||||||
Db 7 TRLTRKRGLK 16

RESULT 10
AAE14541
ID AAE14541 standard; peptide; 34 AA.
XX
AC AAE14541;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human apoB-100 derived peptide p62.
XX
KW Oxidised low density lipoprotein; oxLDL; immunoassay; periaortitis;
KW cardiovascular disease; coronary heart disease; pre-eclampsia;
KW non-insulin-dependent diabetes; endothelial dysfunction; human; apoB-100;
KW peptide p62.
XX
OS Homo sapiens.
XX
PN WO200206314-A2.
XX
PD 24-JAN-2002.
XX
PF 18-JUL-2001; 2001WO-GB003212.
XX
PR 18-JUL-2000; 2000GB-00017641.
XX
PA (ARKT-) ARK THERAPEUTICS LTD.
XX
PI Narvanen O, Yla-Herttuala S;
XX
DR WPI; 2002-179777/23.
XX
PT New peptide useful in enzyme immunoassays for detecting oxidized low
PT density lipoprotein which is a marker of coronary heart disease and other
PT cardiovascular diseases, has affinity for oxidized low density
PT lipoprotein.
XX
PS Claim 6; Page 5; 21pp; English.
XX
CC The invention relates to peptides having affinity for oxidised low
CC density lipoprotein (oxLDL), in cyclised or multimeric form. The peptide
CC is useful in an immunoassay to determine the presence, and optionally,
CC the amount of antibodies in a sample, having affinity for oxLDL.
CC Preferably immobilised peptide is useful for measuring the amount of
CC autoantibodies for oxLDL in a sample, especially a serum or plasma sample
CC from a patient for evaluating the risk of coronary heart diseases, other
CC cardiovascular diseases, and several other disorders such as
CC periaortitis, pre-eclampsia, non-insulin-dependent diabetes and
CC endothelial dysfunction. The peptide of the invention is stable, can be
CC synthesised easily without the need to isolate proteins from a patient's

```
CC   blood, and has a long half-life. The present sequence is human apoB-100
CC   derived peptide p62 used in the invention
XX
SQ   Sequence 34 AA;

  Query Match             89.8%;  Score 44;  DB 5;  Length 34;
  Best Local Similarity   90.0%;  Pred. No. 0.31;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db         25 TRLTRKRGLK 34


RESULT 11
AAW96876
ID   AAW96876 standard; peptide; 36 AA.
XX
AC   AAW96876;
XX
DT   22-APR-1999  (first entry)
XX
DE   Nucleic acid binding domain from apoB-100, residues 3348-3390.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9856938-A1.
XX
PD   17-DEC-1998.
XX
PF   10-JUN-1998;   98WO-US011927.
XX
PR   13-JUN-1997;   97US-00874807.
PR   14-MAY-1998;   98US-00079030.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Guevara JG,  Hoogeveen RC,  Moore JP;
XX
DR   WPI; 1999-070331/06.
XX
PT   Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT   used for delivering nucleic acid to cells for gene therapy and antisense
PT   treatment.
XX
PS   Claim 16; Fig 12C; 293pp; English.
XX
CC   AAW96827-77 represent nucleic acid binding domains derived from human
CC   apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC   of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC   (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC   sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 36 AA;

  Query Match             89.8%;  Score 44;  DB 2;  Length 36;
  Best Local Similarity   90.0%;  Pred. No. 0.33;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db         11 TRLTRKRGLK 20


RESULT 12
AAW64587
ID   AAW64587 standard; peptide; 37 AA.
XX
AC   AAW64587;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Human apolipoprotein peptide fragment #1.
XX
KW   Factor V; human; detection; protein function; blood coagulation; apo;
KW   fat metabolism; Leyden mutation; deep vein thrombosis; apolipoprotein;
KW   Alzheimer's disease; 5,10-methylenetetrahydrofolate reductase; prion;
KW   hypercysteinaemia; factor VII; cardiovascular disease; pathogen; virus.
XX
OS   Homo sapiens.
XX
PN   EP857973-A2.
XX
PD   12-AUG-1998.
XX
PF   12-JAN-1998;   98EP-00890007.
XX
PR   13-JAN-1997;   97AT-00000044.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Moritz B,  Kiessig S,  Lang H,  Schenk V;
XX
DR   WPI; 1998-416142/36.
XX
PT   Detecting or quantifying mutant protein in presence of wild-type protein
PT   comprises reaction with ligand - used to detect mutant blood coagulation
PT   factors or apolipoproteins for diagnosing risk of thrombosis.
XX
PS   Example 2; Page 9; 18pp; German.
XX
CC   AAW64587 and AAW64588 are fragments of human apolipoprotein which are
CC   used with Factor V protein fragments in a novel method to detect the
CC   presence of a mutated protein in a sample that may also contain the
CC   corresponding wild-type protein. The method is used to detect mutations
CC   that alter protein functions (either point mutation or small insertions
CC   or deletions), particularly in proteins involved in blood coagulation or
CC   metabolism of fat. Protein functions which are specially detectable are
CC   the Leyden mutation in factor V (associated with increased risk of deep
CC   vein thrombosis), mutations in apolipoprotein (apo) genes (certain
CC   alleles of apoE indicates increased risk of developing Alzheimer's
CC   disease), thermostable 5,10-methylenetetrahydrofolate reductase
CC   (associated with hypercysteinaemia and venous thrombosis) and factor VII
CC   mutations (associated with increased risk of cardiovascular disease). The
CC   method can also be applied to proteins from pathogens, e.g. viruses or
CC   prions. The method does not require complex apparatus for polymerase
CC   chain reactions, it is simple, standardisable and reliable and is
CC   particularly suited to routine screening. It also allows mutant protein
CC   in a sample to be quantified
XX
SQ   Sequence 37 AA;

  Query Match             89.8%;  Score 44;  DB 2;  Length 37;
  Best Local Similarity   90.0%;  Pred. No. 0.34;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db         11 TRLTRKRGLK 20


RESULT 13
AAW96845
ID   AAW96845 standard; peptide; 51 AA.
```

XX
AC AAW96845;
XX
DT 22-APR-1999 (first entry)
XX
DE Nucleic acid binding domain from apoB-100.
XX
KW Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Claim 16; Page 151; 293pp; English.
XX
CC AAW96827-77 represent nucleic acid binding domains derived from human
CC apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC sequence can be used in the composition of the invention. The
CC specification describes a composition that comprises LDL and
CC apolipoproteins for the binding and in vivo transport of nucleic acids.
CC The composition is used to deliver nucleic acids to eukaryotic cells, in
CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC molecule (or ribozyme). Specifically they are used for gene therapy of
CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC fibrosis and arteriosclerosis
XX
SQ Sequence 51 AA;

Query Match 89.8%; Score 44; DB 2; Length 51;
Best Local Similarity 90.0%; Pred. No. 0.46;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db          6 TRLTRKRGLK 15
```

RESULT 14
ABB37687
ID ABB37687 standard; peptide; 343 AA.
XX
AC ABB37687;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #5193 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
PS Claim 27; SEQ ID NO 30322; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 343 AA;

Query Match 89.8%; Score 44; DB 4; Length 343;
Best Local Similarity 90.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db        169 TRLTRKRGLK 178
```

RESULT 15
ABG52504
ID ABG52504 standard; peptide; 343 AA.
XX
AC ABG52504;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID No 31152.
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 27; SEQ ID NO 31152; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 343 AA;

Query Match 89.8%; Score 44; DB 4; Length 343;
Best Local Similarity 90.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db        169 TRLTRKRGLK 178
```

Search completed: January 13, 2005, 01:43:01
Job time : 78.3798 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:43:16 ; Search time 10.6557 Seconds
(without alignments)
47.947 Million cell updates/sec

Title: US-09-823-418-11
Perfect score: 49
Sequence: 1 TELTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219722 seqs, 51091598 residues

Total number of hits satisfying chosen parameters: 219722

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	42	6	US-10-861-779-1	Sequence 1, Appli
2	44	89.8	3000	8	US-60-636-722-924	Sequence 924, App
3	44	89.8	3262	8	US-60-636-722-923	Sequence 923, App
4	44	89.8	4560	6	US-10-398-200-2	Sequence 2, Appli
5	44	89.8	4563	6	US-10-868-577A-25	Sequence 25, Appl
6	44	89.8	4563	8	US-60-636-722-919	Sequence 919, App
7	44	89.8	4563	8	US-60-636-722-921	Sequence 921, App
8	44	89.8	4563	8	US-60-636-722-922	Sequence 922, App
9	44	89.8	4563	8	US-60-636-722-925	Sequence 925, App
10	41	83.7	720	6	US-10-990-328-11685	Sequence 11685, A
11	41	83.7	720	6	US-10-990-328-11686	Sequence 11686, A
12	41	83.7	720	6	US-10-990-328-11687	Sequence 11687, A
13	41	83.7	720	6	US-10-990-328-11688	Sequence 11688, A
14	41	83.7	720	8	US-60-636-720-278	Sequence 278, App
15	41	83.7	720	8	US-60-636-720-279	Sequence 279, App
16	41	83.7	720	8	US-60-636-720-280	Sequence 280, App
17	41	83.7	720	8	US-60-636-720-281	Sequence 281, App
18	41	83.7	720	8	US-60-636-720-282	Sequence 282, App
19	41	83.7	720	8	US-60-636-720-283	Sequence 283, App
20	34	69.4	1040	6	US-10-777-288A-2245	Sequence 2245, Ap
21	34	69.4	1284	6	US-10-408-765-1008	Sequence 1008, Ap
22	33	67.3	205	6	US-10-732-923-16259	Sequence 16259, A
23	33	67.3	227	6	US-10-732-923-16246	Sequence 16246, A
24	33	67.3	297	1	PCT-US02-09107B-65100	Sequence 65100, A
25	33	67.3	548	1	PCT-US02-09107B-65704	Sequence 65704, A
26	33	67.3	697	6	US-10-999-233-18	Sequence 18, Appl
27	33	67.3	745	6	US-10-999-233-16	Sequence 16, Appl
28	33	67.3	791	6	US-10-999-233-22	Sequence 22, Appl
29	33	67.3	839	6	US-10-999-233-20	Sequence 20, Appl
30	33	67.3	886	6	US-10-732-923-3311	Sequence 3311, Ap
31	33	67.3	963	6	US-10-408-765-2504	Sequence 2504, Ap
32	33	67.3	966	6	US-10-408-765-2155	Sequence 2155, Ap
33	33	67.3	1175	6	US-10-999-233-26	Sequence 26, Appl
34	33	67.3	1175	6	US-10-999-233-30	Sequence 30, Appl
35	33	67.3	1259	6	US-10-999-233-4	Sequence 4, Appli
36	33	67.3	1298	6	US-10-999-233-24	Sequence 24, Appl
37	33	67.3	1298	6	US-10-999-233-28	Sequence 28, Appl
38	33	67.3	1307	6	US-10-999-233-2	Sequence 2, Appli
39	32	65.3	314	6	US-10-990-328-10596	Sequence 10596, A
40	32	65.3	414	6	US-10-990-328-10595	Sequence 10595, A
41	32	65.3	414	6	US-10-990-328-10597	Sequence 10597, A
42	32	65.3	414	8	US-60-631-993-330	Sequence 330, App
43	32	65.3	1380	6	US-10-952-698-164	Sequence 164, App
44	32	65.3	1388	1	PCT-US04-39788-2	Sequence 2, Appli
45	32	65.3	1975	1	PCT-US04-37204-4754	Sequence 4754, Ap

ALIGNMENTS

RESULT 1
US-10-861-779-1
; Sequence 1, Application US/10861779
; GENERAL INFORMATION:
; APPLICANT: Verma, Inder M.
; APPLICANT: Spencer, Brian
; TITLE OF INVENTION: Compositions and Methods For Targeting a
; TITLE OF INVENTION: Polypeptide To the Central Nervous System
; FILE REFERENCE: 66671-131
; CURRENT APPLICATION NUMBER: US/10/861,779
; CURRENT FILING DATE: 2004-06-04
; PRIOR APPLICATION NUMBER: 60/476,482
; PRIOR FILING DATE: 2003-06-05
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-861-779-1

Query Match 89.8%; Score 44; DB 6; Length 42;
Best Local Similarity 90.0%; Pred. No. 0.066;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db         16 TRLTRKRGLK 25
```

RESULT 2
US-60-636-722-924
; Sequence 924, Application US/60636722
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001568
; CURRENT APPLICATION NUMBER: US/60/636,722
; CURRENT FILING DATE: 2004-12-17
; NUMBER OF SEQ ID NOS: 2719
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 924
; LENGTH: 3000
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-636-722-924

Query Match 89.8%; Score 44; DB 8; Length 3000;

```
  Best Local Similarity   90.0%;  Pred. No. 5.3;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       2337 TRLTRKRGLK 2346


RESULT 3
US-60-636-722-923
; Sequence 923, Application US/60636722
; GENERAL INFORMATION:
;  APPLICANT: JOSELOFF, Elizabeth et al.
;  TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001568
;  CURRENT APPLICATION NUMBER: US/60/636,722
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 2719
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 923
;   LENGTH: 3262
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-722-923

  Query Match             89.8%;  Score 44;  DB 8;  Length 3262;
  Best Local Similarity   90.0%;  Pred. No. 5.8;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       2084 TRLTRKRGLK 2093


RESULT 4
US-10-398-200-2
; Sequence 2, Application US/10398200
; GENERAL INFORMATION:
;  APPLICANT: AGNELLO, VINCENT
;  TITLE OF INVENTION: METHOD OF INHIBITING INFECTION BY HCV, OTHER
;  TITLE OF INVENTION:  FLAVIVIRIDAE VIRUSES, AND ANY OTHER VIRUS THAT
;  TITLE OF INVENTION:  COMPLEXES TO LOW DENSITY LIPOPROTEIN OR TO VERY LOW
;  TITLE OF INVENTION:  DENSITY LIPOPRTOEIN IN BLOOD BY PREVENTING VIRAL ENTRY
;  TITLE OF INVENTION:  INTO A CELL
;  FILE REFERENCE: 1513-PCT-00
;  CURRENT APPLICATION NUMBER: US/10/398,200
;  CURRENT FILING DATE:  2003-04-03
;  PRIOR APPLICATION NUMBER: 60/243,594
;  PRIOR FILING DATE: 2000-10-25
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4560
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-398-200-2

  Query Match             89.8%;  Score 44;  DB 6;  Length 4560;
  Best Local Similarity   90.0%;  Pred. No. 8.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3382 TRLTRKRGLK 3391


RESULT 5
US-10-868-577A-25
; Sequence 25, Application US/10868577A
; GENERAL INFORMATION:
;  APPLICANT: Alitalo et al.
;  TITLE OF INVENTION: HEPARIN BINDING VEGFR-3 LIGANDS
;  FILE REFERENCE: 28967/39359A
;  CURRENT APPLICATION NUMBER: US/10/868,577A
;  CURRENT FILING DATE:  2004-06-14
;  PRIOR APPLICATION NUMBER: US 60/478,390
;  PRIOR FILING DATE: 2003-06-12
;  PRIOR APPLICATION NUMBER: US 10/669,176
;  PRIOR FILING DATE: 2003-09-23
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25
;   LENGTH: 4563
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (32)..(126)
;   OTHER INFORMATION: heparin binding domain
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3161)..(3236)
;   OTHER INFORMATION: heparin binding domain
US-10-868-577A-25

  Query Match             89.8%;  Score 44;  DB 6;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 8.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 6
US-60-636-722-919
; Sequence 919, Application US/60636722
; GENERAL INFORMATION:
;  APPLICANT: JOSELOFF, Elizabeth et al.
;  TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001568
;  CURRENT APPLICATION NUMBER: US/60/636,722
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 2719
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 919
;   LENGTH: 4563
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-722-919

  Query Match             89.8%;  Score 44;  DB 8;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 8.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 7
US-60-636-722-921
; Sequence 921, Application US/60636722
; GENERAL INFORMATION:
;  APPLICANT: JOSELOFF, Elizabeth et al.
;  TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001568
;  CURRENT APPLICATION NUMBER: US/60/636,722
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 2719
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 921
;   LENGTH: 4563
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-722-921
```

```
  Query Match             89.8%;  Score 44;  DB 8;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 8.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 8
US-60-636-722-922
; Sequence 922, Application US/60636722
; GENERAL INFORMATION:
;  APPLICANT: JOSELOFF, Elizabeth et al.
;  TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001568
;  CURRENT APPLICATION NUMBER: US/60/636,722
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 2719
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 922
;   LENGTH: 4563
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-722-922

  Query Match             89.8%;  Score 44;  DB 8;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 8.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 9
US-60-636-722-925
; Sequence 925, Application US/60636722
; GENERAL INFORMATION:
;  APPLICANT: JOSELOFF, Elizabeth et al.
;  TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001568
;  CURRENT APPLICATION NUMBER: US/60/636,722
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 2719
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 925
;   LENGTH: 4563
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-722-925

  Query Match             89.8%;  Score 44;  DB 8;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 8.1;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 10
US-10-990-328-11685
; Sequence 11685, Application US/10990328
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele
;  TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION:  ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001495
;  CURRENT APPLICATION NUMBER: US/10/990,328
;  CURRENT FILING DATE:  2004-11-17
;  NUMBER OF SEQ ID NOS: 558824
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11685
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-990-328-11685

  Query Match             83.7%;  Score 41;  DB 6;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 11
US-10-990-328-11686
; Sequence 11686, Application US/10990328
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele
;  TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION:  ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001495
;  CURRENT APPLICATION NUMBER: US/10/990,328
;  CURRENT FILING DATE:  2004-11-17
;  NUMBER OF SEQ ID NOS: 558824
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11686
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-990-328-11686

  Query Match             83.7%;  Score 41;  DB 6;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 12
US-10-990-328-11687
; Sequence 11687, Application US/10990328
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele
;  TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION:  ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001495
;  CURRENT APPLICATION NUMBER: US/10/990,328
;  CURRENT FILING DATE:  2004-11-17
;  NUMBER OF SEQ ID NOS: 558824
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11687
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-990-328-11687

  Query Match             83.7%;  Score 41;  DB 6;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642
```

```
RESULT 13
US-10-990-328-11688
; Sequence 11688, Application US/10990328
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele
;  TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION:  ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001495
;  CURRENT APPLICATION NUMBER: US/10/990,328
;  CURRENT FILING DATE:  2004-11-17
;  NUMBER OF SEQ ID NOS: 558824
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11688
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-990-328-11688

  Query Match             83.7%;  Score 41;  DB 6;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 14
US-60-636-720-278
; Sequence 278, Application US/60636720
; GENERAL INFORMATION:
;  APPLICANT: DOMON, Bruno et al.
;  TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001567
;  CURRENT APPLICATION NUMBER: US/60/636,720
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 4325
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 278
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-720-278

  Query Match             83.7%;  Score 41;  DB 8;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 15
US-60-636-720-279
; Sequence 279, Application US/60636720
; GENERAL INFORMATION:
;  APPLICANT: DOMON, Bruno et al.
;  TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
;  FILE REFERENCE: CL001567
;  CURRENT APPLICATION NUMBER: US/60/636,720
;  CURRENT FILING DATE:  2004-12-17
;  NUMBER OF SEQ ID NOS: 4325
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 279
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-636-720-279

  Query Match             83.7%;  Score 41;  DB 8;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


Search completed: January 13, 2005, 02:20:03
Job time : 10.6557 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:30:05 ; Search time 14.4262 Seconds
(without alignments)
66.696 Million cell updates/sec

Title: US-09-823-418-11
Perfect score: 49
Sequence: 1 TELTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	275	2	E60950	apolipoprotein B-1
2	44	89.8	596	2	S32802	apolipoprotein B -
3	44	89.8	4563	1	LPHUB	apolipoprotein B-1
4	40	81.6	269	2	C60950	apolipoprotein B-1
5	40	81.6	274	2	A60950	apolipoprotein B-1
6	40	81.6	779	2	JH0102	apolipoprotein B -
7	37	75.5	309	1	E65112	hypothetical 34.6
8	37	75.5	309	2	E85985	hypothetical prote
9	37	75.5	309	2	B91140	hypothetical prote
10	36	73.5	254	2	H95070	hypothetical prote
11	36	73.5	264	2	F97938	hypothetical prote
12	35	71.4	149	2	E72338	conserved hypothet
13	35	71.4	214	2	AE2934	transcription regu
14	35	71.4	219	2	B98348	probable transcrip
15	35	71.4	1112	2	T47784	hypothetical prote
16	34	69.4	550	1	WZCKI	isocitrate lyase (
17	34	69.4	784	2	JH0101	apolipoprotein B-1
18	34	69.4	1058	2	S65460	apolipoprotein B -
19	34	69.4	1429	2	T41699	C2-domain family p
20	34	69.4	1778	2	JT0382	apolipoprotein B -
21	34	69.4	2629	2	I46569	apolipoprotein B -
22	33	67.3	219	2	A53305	pentose-5-phosphat
23	33	67.3	227	2	H84614	probable MADS-box
24	33	67.3	235	2	E86826	amino acid permeas
25	33	67.3	240	2	A70463	rRNA methylase - A
26	33	67.3	249	2	T16924	hypothetical prote
27	33	67.3	309	2	AH0906	conserved hypothet
28	33	67.3	443	2	D72383	NADH oxidase - The
29	33	67.3	548	2	G81959	conserved hypothet
30	33	67.3	583	2	T04531	nine-cis-epoxycaro
31	33	67.3	642	2	T05683	hypothetical prote
32	33	67.3	963	2	T26022	hypothetical prote
33	33	67.3	1091	2	T35822	probable regulator
34	33	67.3	1253	1	A44400	myosin heavy chain
35	33	67.3	1254	2	A54818	myosin-VI [similar
36	33	67.3	1265	2	A59299	unconventional myo
37	33	67.3	1407	2	S59823	probable membrane
38	33	67.3	1615	2	JC6510	ras-responsive ele
39	33	67.3	2044	2	T13704	still life protein
40	33	67.3	2064	2	T13707	still life protein
41	32	65.3	107	2	S12607	salivary glue prot
42	32	65.3	112	2	S33822	salivary glue prot
43	32	65.3	172	2	F69506	probable 2-oxoisov
44	32	65.3	183	2	E84119	ATP synthase delta
45	32	65.3	237	2	T19702	hypothetical prote

ALIGNMENTS

RESULT 1
E60950
apolipoprotein B-100 - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: E60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LD
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: E60950
A;Molecule type: mRNA
A;Residues: 1-275 <LAW>
A;Cross-references: UNIPROT:Q7LZ77
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein

Query Match 89.8%; Score 44; DB 2; Length 275;
Best Local Similarity 90.0%; Pred. No. 0.49;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db        221 TSLTRKRGLK 230
```

RESULT 2
S32802
apolipoprotein B - crab-eating macaque (fragment)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S32802
R;Pape, M.E.; Castle, C.K.; Murray, R.W.; Funk, G.M.; Hunt, C.E.; Marotti, K.R.; Melchi
Biochim. Biophys. Acta 1086, 326-334, 1991
A;Title: Apo B metabolism in the cynomolgus monkey: evidence for post-transcriptional r
A;Reference number: S32802; MUID:92075708; PMID:1742325
A;Accession: S32802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-596 <PAP>
A;Cross-references: UNIPROT:Q28473; EMBL:X15737; NID:g38047; PIDN:CAA33755.1; PID:g9301
C;Superfamily: apolipoprotein B

Query Match 89.8%; Score 44; DB 2; Length 596;
Best Local Similarity 90.0%; Pred. No. 0.98;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db        226 TRLTRKRGLK 235
```

RESULT 3
LPHUB
apolipoprotein B-100 precursor - human
N;Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
C;Species: Homo sapiens (man)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: A27850; A25679; A25263; A25267; A25266; A24320; A24684; A23817; A25774; A26
4452; I61909; I59510; I39474; I39469; I84624; I37179; PS0058
R;Ludwig, E.H.; Blackhart, B.D.; Pierotti, V.R.; Caiati, L.; Fortier, C.; Knott, T.; Scc
DNA 6, 363-372, 1987
A;Title: DNA sequence of the human apolipoprotein B gene.
A;Reference number: A27850; MUID:88003974; PMID:3652907
A;Accession: A27850
A;Molecule type: DNA
A;Residues: 1-617,'A',619-1929,'F',1931-3318,'D',3320-3426,'T',3428-3431,'Q',3433-3731,'
A;Cross-references: UNIPROT:P04114; UNIPROT:P78482; UNIPROT:P78479; UNIPROT:Q9UMN0; UNIE
R;Cladaras, C.; Hadzopoulou-Cladaras, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A;Title: The complete sequence and structural analysis of human apolipoprotein B-100: re
A;Reference number: A91058; MUID:87161758; PMID:3030729
A;Accession: A25679
A;Molecule type: mRNA
A;Residues: 1-11,15-2539,'S',2541-3823,'R',3825-4563 <CLA>
A;Note: 1109-Asp was also found
R;Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McCa
Nucleic Acids Res. 14, 7501-7503, 1986
A;Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A;Reference number: A93639; MUID:87016385; PMID:3763409
A;Accession: A25263
A;Molecule type: mRNA
A;Residues: 1-272,'N',274-617,'A',619-1217,'E',1219-2091,'V',2093-2364,'T',2366-2679,'Q'
A;Cross-references: GB:X04506; NID:g34330; PIDN:CAA28191.1; PID:g34331
R;Law, S.W.; Grant, S.M.; Higuchi, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer Jr
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A;Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A;Reference number: A94134; MUID:87041416; PMID:3464946
A;Accession: A25267
A;Molecule type: mRNA
A;Residues: 1-617,'A',619-703,'P',705-792,'R',794-1270,'S',1272-1866,'G',1868-2036,'N',2
4189-4220,'M',4222-4563 <LAW>
A;Note: the codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4188, and
R;Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Gotto Jr., A.M
J. Biol. Chem. 261, 12918-12921, 1986
A;Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A;Reference number: A92556; MUID:87008488; PMID:3759943
A;Accession: A25266
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-328,'V',330-644,'I',646-918,'P',920-3318,'D',3320-3426,'T',3428-
9-4132,'G',4134-4180,'E',4182-4563 <CHE>
A;Cross-references: GB:J02610; NID:g178803; PIDN:AAA35549.1; PID:g178804
A;Note: a total of 2366 residues were confirmed by direct sequencing of tryptic peptides
R;Protter, A.A.; Hardman, D.A.; Sato, K.Y.; Schilling, J.W.; Yamanaka, M.; Hort, Y.J.; H
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A;Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein
A;Reference number: A24320; MUID:86287319; PMID:3461454
A;Accession: A24320
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-617,'A',619-941,'YYIWSLPPKP',951-1138,'PTGRLPNCFSNGLICYSLWLHSFQE
A;Cross-references: GB:M14081; NID:g178795; PIDN:AAA51752.1; PID:g553189
R;Law, S.W.; Lackner, K.J.; Hospattankar, A.V.; Anchors, J.M.; Sakaguchi, A.Y.; Naylor,
Proc. Natl. Acad. Sci. U.S.A. 82, 8340-8344, 1985
A;Title: Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of
A;Reference number: A24684; MUID:86094221; PMID:3001697
A;Accession: A24684
A;Molecule type: mRNA
A;Residues: 485-617,'A',619-1044 <LA2>
A;Cross-references: GB:M12480; NID:g178791; PIDN:AAA51751.1; PID:g178792
R;Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; Ki
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A;Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipopr
A;Reference number: A94088; MUID:86149325; PMID:3513177
A;Accession: A23817
A;Molecule type: mRNA
A;Residues: 1-291 <PRO>
A;Cross-references: GB:M12681; NID:g178797; PIDN:AAA51753.1; PID:g178798
R;Deeb, S.S.; Motulsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A;Title: A partial cDNA clone for human apolipoprotein B.
A;Reference number: A25774; MUID:85270450; PMID:3860836
A;Accession: A25774
A;Molecule type: mRNA
A;Residues: 709-791,'SSSWKAASHGCPHSAGD',810-906 <DEE>
A;Cross-references: GB:K03175; NID:g178821; PIDN:AAA51759.1; PID:g178822
R;Carlsson, P.; Darnfors, C.; Olofsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A;Title: Analysis of the human apolipoprotein B gene; complete structure of the B-74 re
A;Reference number: A91565; MUID:87191999; PMID:2883086
A;Accession: A26533
A;Molecule type: mRNA
A;Residues: 1282-2721,2742-3290,'L',3292-3336,'N',3338-3948,'F',3950-3963,'Y',3965-4180,
A;Cross-references: GB:M15421; NID:g178817; PIDN:AAA51758.1; PID:g178818
R;Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Sato, K.Y.; Lau, K.; Yamani
Biochemistry 26, 5478-5486, 1987
A;Title: Structural comparison of human apolipoproteins B-48 and B-100.
A;Reference number: A29671; MUID:88050832; PMID:3676265
A;Accession: A29671
A;Molecule type: mRNA
A;Residues: 1671-2323,'PYW',2327-2352,'H',2354-2398 <HAR>
A;Cross-references: GB:M17367; NID:g178731; PIDN:AAA51741.1; PID:g178732
R;Shoulders, C.C.; Myant, N.B.; Sidoli, A.; Rodriguez, J.C.; Cortese, C.; Baralle, F.E.,
Atherosclerosis 58, 277-289, 1985
A;Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for more than o
A;Reference number: A90084; MUID:86130855; PMID:3841481
A;Accession: A29287
A;Molecule type: mRNA
A;Residues: 3846-4298 <SHO>
R;Pfitzner, R.; Wagener, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986
A;Title: Isolation, expression and characterization of a human apolipoprotein B 100-spe
A;Reference number: A25572; MUID:87076044; PMID:3024665
A;Accession: A25572
A;Molecule type: mRNA
A;Residues: 4219-4337,'S',4339-4563 <PFI>
A;Cross-references: GB:M36676
R;Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow, J.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269, 1985
A;Reference number: A24738; MUID:86042646; PMID:2932736
A;Accession: A24738
A;Molecule type: mRNA
A;Residues: 'N',3729-3731,'I',3733-3875,'A',3877-3948,'F',3950-3963,'Y',3965-3982,'S',3
A;Cross-references: GB:M12413; NID:g178735; PIDN:AAA51742.1; PID:g178736
R;Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.; Silberman, S.R.; Ca
Science 238, 363-366, 1987
A;Title: Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific i
A;Reference number: A40133; MUID:88018019; PMID:3659919
A;Accession: B40133
A;Molecule type: mRNA
A;Residues: 2165-2179 <CH1>
A;Cross-references: GB:M18036; NID:g178799; PIDN:AAA51754.1; PID:g178800
A;Note: this mRNA includes the stop codon of the organ-specific mRNA for apo48
A;Accession: A40133
A;Molecule type: protein
A;Residues: 51-75;101-110;129-139;158-174;197-207;276-287;298-304;306-314;526-532;538-5
36;1486-1498;1537-1556;1563-1572;1601-1610;1647-1661;1697-1724;1770-1781;1859-1897;1968
A;Note: these fragments were derived from apo48
R;Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987
A;Title: Carboxyl terminal analysis of human B-48 protein confirms the novel mechanism
A;Reference number: A28002; MUID:88106542; PMID:3426612
A;Accession: A28002
A;Molecule type: mRNA
A;Residues: 2129-2179,2181-2235 <HA2>
A;Cross-references: GB:M18471
A;Experimental source: intestine
A;Note: this mRNA from intestine includes a stop codon created by RNA editing in place
R;Mehrabian, M.; Schumaker, V.N.; Fareed, G.C.; West, R.; Johnson, D.F.; Kirchgessner,

Nucleic Acids Res. 13, 6937-6953, 1985
A;Title: Human apolipoprotein B: identification of cDNA clones and characterization of m
A;Reference number: A24269; MUID:86041888; PMID:3903660
A;Accession: A24269
A;Molecule type: mRNA
A;Residues: 3056-3159 <MEH>
A;Cross-references: GB:X03045; NID:g28783; PIDN:CAA26850.1; PID:g929609
R;Hospattankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer Jr., H.B.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A;Title: Identification of a novel in-frame translational stop codon in human intestine
A;Reference number: A29659; MUID:88049670; PMID:2445342
A;Accession: A29659
A;Molecule type: mRNA
A;Residues: 2169-2179 <HOS>
A;Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A;Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human intest
ch encodes the 250K apoB-48, CAA encoding 2180-Gln is substituted by the stop codon TAA,
R;Yang, C.; Kim, T.W.; Weng, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
A;Title: Isolation and characterization of sulfhydryl and disulfide peptides of human ap
A;Reference number: A35783; MUID:90319144; PMID:2115173
A;Contents: disulfide bonds
A;Accession: A35783
A;Molecule type: protein
A;Residues: 28-41;76-97,'I',99-100;175-193;206-215;239-249;259-266;357-399;455-490;512-5

A;Note: cysteines at positions 1112, 1422, 1505, 1662, 3761, 3917, and 4217 have free su
R;LeBoeuf, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
FEBS Lett. 170, 105-108, 1984
A;Title: Human apolipoprotein B: partial amino acid sequence.
A;Reference number: A22006; MUID:84208786; PMID:6373369
A;Accession: A22006
A;Molecule type: protein
A;Residues: 873-892,'K',894-896 <LE1>
A;Accession: B22006
A;Molecule type: protein
A;Residues: 3113,'L',3115-3130,'R',3132-3133,'P',3135-3136,'R' <LE2>
R;Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Caiati, L.; Onasch, M.A.; Wallis, S.C.;
J. Biol. Chem. 261, 15364-15367, 1986
A;Title: Structure of the human apolipoprotein B gene.
A;Reference number: A92564; MUID:87057153; PMID:2946672
A;Contents: annotation; gene structure
R;Wagener, R.; Pfitzner, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
A;Title: Studies on the organization of the human apolipoprotein B 100 gene.
A;Reference number: A90715; MUID:87271140; PMID:2886136
A;Contents: annotation; gene structure
R;Weisgraber, K.H.; Rall Jr., S.C.
J. Biol. Chem. 262, 11097-11103, 1987
A;Title: Human apolipoprotein B-100 heparin-binding sites.
A;Reference number: A92605; MUID:87280197; PMID:3301850
A;Contents: annotation; heparin binding and disulfide bond
R;Dashti, N.; Lee, D.M.; Mok, T.
Biochem. Biophys. Res. Commun. 137, 493-499, 1986
A;Title: Apolipoprotein B is a calcium binding protein.
A;Reference number: A90125; MUID:86242245; PMID:3087360
A;Contents: annotation; calcium binding
R;Carlsson, P.; Olofsson, S.O.; Bondjers, G.; Darnfors, C.; Wiklund, O.; Bjursell, G.
Nucleic Acids Res. 13, 8813-8826, 1985
A;Title: Molecular cloning of human apolipoprotein B cDNA.
A;Reference number: I37178; MUID:86093680; PMID:3841204
A;Accession: I37180

  Query Match 89.8%; Score 44; DB 1; Length 4563;
  Best Local Similarity 90.0%; Pred. No. 6.3;
  Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
  | ||||||||
Db 3385 TRLTRKRGLK 3394

RESULT 4
C60950
apolipoprotein B-100 - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: C60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LD
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: C60950
A;Molecule type: DNA
A;Residues: 1-269 <LAW>
A;Cross-references: UNIPROT:Q60537; UNIPROT:Q60536
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein

  Query Match 81.6%; Score 40; DB 2; Length 269;
  Best Local Similarity 80.0%; Pred. No. 3;
  Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
  : ||||||||
Db 216 SRLTRKRGLK 225

RESULT 5
A60950
apolipoprotein B-100 - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Dec-1993 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004
C;Accession: A60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LD
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: A60950
A;Molecule type: mRNA
A;Residues: 1-274 <LAW>
A;Cross-references: UNIPROT:Q7M2U9
A;Note: authors translated the codon GAT for residue 155 as His
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein

  Query Match 81.6%; Score 40; DB 2; Length 274;
  Best Local Similarity 80.0%; Pred. No. 3;
  Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TELTRKRGLK 10
  : ||||||||
Db 221 SSLTRKRGLK 230

RESULT 6
JH0102
apolipoprotein B - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: JH0102
R;Smith, T.J.
submitted to GenBank, June 1990
A;Reference number: A38864
A;Accession: JH0102
A;Molecule type: DNA
A;Residues: 1-779 <SMI>
A;Cross-references: UNIPROT:Q60536; GB:M35187
A;Note: this is a revision to the sequence from reference JH0101
R;Smith, T.J.; Hautamaa, D.; Maeda, N.
Gene 87, 309-310, 1990
A;Title: Sequence of the putative low-density lipoprotein receptor-binding regions of a
A;Reference number: JH0101; MUID:90236327; PMID:2332175
A;Contents: annotation
A;Note: this sequence has been revised in reference A38864
C;Genetics:

A;Gene: apoB
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;
F;435-445/Region: receptor binding
F;646-656/Region: receptor binding

Query Match 81.6%; Score 40; DB 2; Length 779;
Best Local Similarity 80.0%; Pred. No. 7.8;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              : ||||||||
Db        642 SRLTRKRGLK 651
```

RESULT 7
E65112
hypothetical 34.6 kD protein in arcB-gltB intergenic region - Escherichia coli (strain K
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: E65112
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: E65112
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-309 <BLAT>
A;Cross-references: UNIPROT:P45476; GB:AE000400; GB:U00096; NID:g2367203; PIDN:AAC76243.
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yhcC
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0486

Query Match 75.5%; Score 37; DB 1; Length 309;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              |:| |:||||
Db        170 TQLARQRGLK 179
```

RESULT 8
E85985
hypothetical protein yhcC [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: E85985
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85985
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <STO>
A;Cross-references: UNIPROT:P45476; GB:AE005174; NID:g12517832; PIDN:AAG58345.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yhcC
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0486

Query Match 75.5%; Score 37; DB 2; Length 309;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              |:| |:||||
Db        170 TQLARQRGLK 179
```

RESULT 9
B91140
hypothetical protein ECs4090 [imported] - Escherichia coli (strain O157:H7, substrain R
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: B91140
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B91140
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <HAY>
A;Cross-references: UNIPROT:P45476; GB:BA000007; PIDN:BAB37513.1; PID:g13363563; GSPDB:(
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4090
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0486

Query Match 75.5%; Score 37; DB 2; Length 309;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              |:| |:||||
Db        170 TQLARQRGLK 179
```

RESULT 10
H95070
hypothetical protein SP0609 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: H95070
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: H95070
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <KUR>
A;Cross-references: UNIPROT:Q97S14; GB:AE005672; PIDN:AAK74761.1; PID:g14972084; GSPDB:(
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0609

Query Match 73.5%; Score 36; DB 2; Length 254;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              ||| :|:|||
Db        149 TELGKKKGLK 158
```

RESULT 11
F97938
hypothetical protein glnH [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: F97938
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001

```
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: F97938
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-264 <KUR>
A;Cross-references: UNIPROT:Q8CWT1; GB:AE007317; PIDN:AAK99338.1; PID:g15458109; GSPDB:G
C;Genetics:
A;Gene: glnH

  Query Match             73.5%;  Score 36;  DB 2;  Length 264;
  Best Local Similarity   70.0%;  Pred. No. 18;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| :|:|||
Db        159 TELGKKKGLK 168


RESULT 12
E72338
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: E72338
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: E72338
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-149 <ARN>
A;Cross-references: UNIPROT:Q9WZI9; GB:AE001744; GB:AE000512; NID:g4981254; PIDN:AAD3581
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0730
C;Superfamily: conserved hypothetical protein HI0670

  Query Match             71.4%;  Score 35;  DB 2;  Length 149;
  Best Local Similarity   77.8%;  Pred. No. 17;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 ELTRKRGLK 10
              || ||:|||
Db        111 ELLRKKGLK 119


RESULT 13
AE2934
transcription regulator, TetR family Atu3075 [imported] - Agrobacterium tumefaciens (str
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C;Accession: AE2934
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE2934
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <KUR>
A;Cross-references: UNIPROT:Q8UBE1; GB:AE008689; PIDN:AAL43891.1; PID:g17741439; GSPDB:G
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3075
A;Map position: linear chromosome

  Query Match             71.4%;  Score 35;  DB 2;  Length 214;
  Best Local Similarity   87.5%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRG 8
              || |||||
Db          2 TETTRKRG 9


RESULT 14
B98348
probable transcription regulator PA2931 [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: B98348
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B98348
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <KUR>
A;Cross-references: UNIPROT:Q8UBE1; GB:AE007870; PIDN:AAK90308.1; PID:g15160337; GSPDB:
C;Genetics:
A;Gene: AGR_L_3469
A;Map position: linear chromosome

  Query Match             71.4%;  Score 35;  DB 2;  Length 219;
  Best Local Similarity   87.5%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRG 8
              || |||||
Db          7 TETTRKRG 14


RESULT 15
T47784
hypothetical protein F17J16.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47784
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24476
A;Accession: T47784
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1112 <DAN>
A;Cross-references: UNIPROT:Q9LYT4; EMBL:AL163527
A;Experimental source: cultivar Columbia; BAC clone F17J16
C;Genetics:
A;Map position: 3
A;Introns: 31/3; 74/1; 103/2; 153/2; 193/2; 247/3; 309/2; 367/2; 416/1; 457/2; 517/3; 5
A;Note: F17J16.70

  Query Match             71.4%;  Score 35;  DB 2;  Length 1112;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRG 8
              ||| ||||
Db        433 TELVRKRG 440


Search completed: January 13, 2005, 01:52:38
Job time : 16.4262 secs
```

THIS PAGE IS BLANK

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:15:45 ; Search time 78.0328 Seconds
(without alignments)
73.735 Million cell updates/sec

Title: US-09-823-418-11
Perfect score: 49
Sequence: 1 TELTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	89.8	275	2	Q7LZ77	Q7lz77 gallus gall
2	44	89.8	387	2	Q7YQN2	Q7yqn2 phalanger o
3	44	89.8	400	2	Q7YQM9	Q7yqm9 ornithorhyn
4	44	89.8	405	2	Q7YQN0	Q7yqn0 tachyglossu
5	44	89.8	414	2	Q7YQR5	Q7yqr5 aotus vocif
6	44	89.8	596	2	Q28473	Q28473 macaca fasc
7	44	89.8	3262	2	Q13788	Q13788 homo sapien
8	44	89.8	4563	1	APB_HUMAN	P04114 homo sapien
9	44	89.8	4563	2	Q7Z600	Q7z600 homo sapien
10	41	83.7	164	2	Q6PIM8	Q6pim8 homo sapien
11	41	83.7	164	2	AAH32144	Aah32144 homo sapi
12	41	83.7	720	1	LCF3_HUMAN	O95573 homo sapien
13	41	83.7	720	2	Q8IUM9	Q8ium9 homo sapien
14	41	83.7	720	2	BAB72139	Bab72139 homo sapi
15	40	81.6	263	2	Q7YQQ0	Q7yqq0 procavia ca
16	40	81.6	274	2	Q7M2U9	Q7m2u9 oryctolagus
17	40	81.6	304	2	Q7YQP9	Q7yqp9 echinops te
18	40	81.6	314	2	Q7YQN8	Q7yqn8 ictonyx str
19	40	81.6	316	2	Q7YQP3	Q7yqp3 nandinia bi
20	40	81.6	318	2	Q7YQN9	Q7yqn9 zalophus ca
21	40	81.6	319	2	Q7YQP0	Q7yqp0 vulpes vulp
22	40	81.6	319	2	Q7YQP2	Q7yqp2 panthera le
23	40	81.6	320	2	Q7YQP4	Q7yqp4 manis sp. k
24	40	81.6	322	2	Q7YQP5	Q7yqp5 manis sp. k
25	40	81.6	339	2	Q7YR05	Q7yr05 macroscelid
26	40	81.6	361	2	Q7YQP8	Q7yqp8 amblysomus
27	40	81.6	364	2	Q7YQQ1	Q7yqq1 dugong dugo
28	40	81.6	386	2	Q7YQR1	Q7yqr1 tupaia tana
29	40	81.6	392	2	Q7YR11	Q7yr11 tarsius syr
30	40	81.6	411	2	Q7YQP7	Q7yqp7 ochotona pr
31	40	81.6	421	2	Q7TN68	Q7tn68 glaucomys v
32	40	81.6	422	2	Q7YR12	Q7yr12 talpa europ
33	40	81.6	423	2	Q7YQQ9	Q7yqq9 sorex monti
34	40	81.6	426	2	Q7YQR2	Q7yqr2 alces alces
35	40	81.6	429	2	Q7YQQ8	Q7yqq8 crocidura f
36	40	81.6	432	2	Q7YR10	Q7yr10 diceros bic
37	40	81.6	436	2	Q7YQM8	Q7yqm8 nyctimene a
38	40	81.6	438	2	Q7YQM7	Q7yqm7 pteropus hy
39	40	81.6	438	2	Q7YQR4	Q7yqr4 balaena mys
40	40	81.6	438	2	Q7YR04	Q7yr04 rousettus a
41	40	81.6	441	2	Q7YQR3	Q7yqr3 phocoenoide
42	40	81.6	443	2	Q7YQN5	Q7yqn5 megaderma l
43	40	81.6	443	2	Q7YQP6	Q7yqp6 lepus ameri
44	40	81.6	445	2	Q7YQN6	Q7yqn6 bradypus tr
45	40	81.6	445	2	Q7YQQ7	Q7yqq7 tapirus bai

ALIGNMENTS

RESULT 1
Q7LZ77
ID Q7LZ77 PRELIMINARY; PRT; 275 AA.
AC Q7LZ77;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Apolipoprotein B-100 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90324804; PubMed=2373961;
RA Law A., Scott J.;
RT "A cross-species comparison of the apolipoprotein B domain that binds
RT to the LDL receptor.";
RL J. Lipid Res. 31:1109-1120(1990).
DR PIR; E60950; E60950.
FT NON_TER 1 1
FT NON_TER 275 275
SQ SEQUENCE 275 AA; 30678 MW; B7D8DA054E04B255 CRC64;

Query Match 89.8%; Score 44; DB 2; Length 275;
Best Local Similarity 90.0%; Pred. No. 1.8;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TELTRKRGLK 10
              | ||||||||
Db        221 TSLTRKRGLK 230
```

RESULT 2
Q7YQN2
ID Q7YQN2 PRELIMINARY; PRT; 387 AA.
AC Q7YQN2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Apolipoprotein B 100 (Fragment).
GN Name=apoB-100;
OS Phalanger orientalis (gray cuscus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Phalanger.
OX NCBI_TaxID=42473;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22761261; PubMed=12878460;
RA Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT "A new phylogenetic marker, apolipoprotein B, provides compelling
RT evidence for eutherian relationships.";
RL Mol. Phylogenet. Evol. 28:225-240(2003).

```
DR   EMBL; AF548431; AAP97387.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     387    387
SQ   SEQUENCE   387 AA;  43230 MW;  8300A9D7C54B42B0 CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 387;
  Best Local Similarity   90.0%;  Pred. No. 2.5;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db        260 TSLTRKRGLK 269


RESULT 3
Q7YQM9
ID   Q7YQM9      PRELIMINARY;      PRT;   400 AA.
AC   Q7YQM9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Ornithorhynchus anatinus (Duckbill platypus).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
OX   NCBI_TaxID=9258;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548434; AAP97390.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     400    400
SQ   SEQUENCE   400 AA;  44611 MW;  DC79873CA6D01CFA CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 400;
  Best Local Similarity   90.0%;  Pred. No. 2.6;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db        260 TSLTRKRGLK 269


RESULT 4
Q7YQN0
ID   Q7YQN0      PRELIMINARY;      PRT;   405 AA.
AC   Q7YQN0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Tachyglossus aculeatus (Australian echidna).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Monotremata; Tachyglossidae; Tachyglossus.
OX   NCBI_TaxID=9261;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548433; AAP97389.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     405    405
SQ   SEQUENCE   405 AA;  44975 MW;  551A98557E8B081D CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 405;
  Best Local Similarity   90.0%;  Pred. No. 2.6;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db        260 TSLTRKRGLK 269


RESULT 5
Q7YQR5
ID   Q7YQR5      PRELIMINARY;      PRT;   414 AA.
AC   Q7YQR5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Aotus vociferans (Spix's owl monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX   NCBI_TaxID=57176;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548396; AAP97352.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     414    414
SQ   SEQUENCE   414 AA;  45955 MW;  EEFA8492157E1BDE CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 414;
  Best Local Similarity   90.0%;  Pred. No. 2.7;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db        258 TRLTRKRGLK 267


RESULT 6
Q28473
ID   Q28473      PRELIMINARY;      PRT;   596 AA.
AC   Q28473;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Apolipoprotein B (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92075708; PubMed=1742325;
RA   Pape M.E., Castle C.K., Murray R.W., Funk G.M., Hunt C.E.,
RA   Marotti K.R., Melchior G.W.;
RT   "Apo B metabolism in the cynomolgus monkey: evidence for post-
RT   transcriptional regulation.";
RL   Biochim. Biophys. Acta 1086:326-334(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Murray R.;
```

```
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X15737; CAA33755.1; -.
DR   PIR; S32802; S32802.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     596    596
SQ   SEQUENCE   596 AA;  66757 MW;  B13BBA74E25C3120 CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 596;
  Best Local Similarity   90.0%;  Pred. No. 4;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db        226 TRLTRKRGLK 235


RESULT 7
Q13788
ID   Q13788      PRELIMINARY;      PRT;  3262 AA.
AC   Q13788;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   APOB protein (Fragment).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87191999; PubMed=2883086;
RA   Carlsson P., Darnfors C., Olofsson S.O., Bjursell G.;
RT   "Analysis of the human apolipoprotein B gene; complete structure of
RT   the B-74 region.";
RL   Gene 49:29-51(1986).
DR   EMBL; M15421; AAA51758.1; -.
DR   PIR; A27850; LPHUB.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0005319; F:lipid transporter activity; NAS.
DR   GO; GO:0006869; P:lipid transport; NAS.
FT   NON_TER       1      1
SQ   SEQUENCE   3262 AA;  370140 MW;  56603BC0618DD40D CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 3262;
  Best Local Similarity   90.0%;  Pred. No. 24;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       2084 TRLTRKRGLK 2093


RESULT 8
APB_HUMAN
ID   APB_HUMAN      STANDARD;      PRT;  4563 AA.
AC   P04114; O00502; Q13787;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
DE   B-48 (Apo B-48)].
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87016385; PubMed=3763409;
RA   Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
RA   Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
RT   "Complete cDNA and derived protein sequence of human apolipoprotein B-
RT   100.";
RL   Nucleic Acids Res. 14:7501-7503(1986).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT GLU-4181.
RX   MEDLINE=88003974; PubMed=3652907;
RA   Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
RA   Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
RT   "DNA sequence of the human apolipoprotein B gene.";
RL   DNA 6:363-372(1987).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS ILE-98 AND GLU-4181.
RX   MEDLINE=87008488; PubMed=3759943;
RA   Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
RA   Gotto A.M. Jr., Chan L.;
RT   "The complete cDNA and amino acid sequence of human apolipoprotein B-
RT   100.";
RL   J. Biol. Chem. 261:12918-12921(1986).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87041416; PubMed=3464946;
RA   Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
RA   Lee N., Brewer H.B. Jr.;
RT   "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
RT   derived amino acid sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87161758; PubMed=3030729;
RA   Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
RA   Zannis V.I.;
RT   "The complete sequence and structural analysis of human apolipoprotein
RT   B-100: relationship between apoB-100 and apoB-48 forms.";
RL   EMBO J. 5:3495-3507(1986).
RN   [6]
RP   SEQUENCE OF 709-906 FROM N.A.
RX   MEDLINE=85270450; PubMed=3860836;
RA   Deeb S.S., Motulsky A.G., Albers J.J.;
RT   "A partial cDNA clone for human apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN   [7]
RP   SEQUENCE OF 3056-3159 FROM N.A.
RX   MEDLINE=86041888; PubMed=3903660;
RA   Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
RA   Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
RT   "Human apolipoprotein B: identification of cDNA clones and
RT   characterization of mRNA.";
RL   Nucleic Acids Res. 13:6937-6953(1985).
RN   [8]
RP   SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A., AND VARIANT GLU-4181.
RX   MEDLINE=86093680; PubMed=3841204;
RA   Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
RA   Bjursell G.;
RT   "Molecular cloning of human apolipoprotein B cDNA.";
RL   Nucleic Acids Res. 13:8813-8826(1985).
RN   [9]
RP   SEQUENCE OF 3109-4563 FROM N.A.
RX   MEDLINE=85300528; PubMed=2994225;
RA   Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Urdea M.S.,
RA   Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M.,
RA   Priestley L.M., Robertson E., Rall L.B., Betsholtz C., Shows T.B.,
RA   Mahley R.W., Scott J.;
RT   "Human apolipoprotein B: structure of carboxyl-terminal domains, sites
RT   of gene expression, and chromosomal localization.";
RL   Science 230:37-43(1985).
RN   [10]
RP   SEQUENCE OF 1-291 FROM N.A.
RX   MEDLINE=86149325; PubMed=3513177;
RA   Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA   Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
RT   "Isolation of a cDNA clone encoding the amino-terminal region of human
RT   apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
```

```
RN   [11]
RP   SEQUENCE OF 1-1670 FROM N.A., AND VARIANT ILE-98.
RX   MEDLINE=86287319; PubMed=3461454;
RA   Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanaka M.,
RA   Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT   "Analysis of cDNA clones encoding the entire B-26 region of human
RT   apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN   [12]
RP   PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX   MEDLINE=88018019; PubMed=3659919;
RA   Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA   Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
RA   Gotto A.M. Jr., Li W.-H., Chan L.;
RT   "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT   specific in-frame stop codon.";
RL   Science 238:363-366(1987).
RN   [13]
RP   DOMAINS.
RX   MEDLINE=87039351; PubMed=3773997;
RA   Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA   Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA   Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA   Levy-Wilson B., Scott J.;
RT   "Complete protein sequence and identification of structural domains of
RT   human apolipoprotein B.";
RL   Nature 323:734-738(1986).
RN   [14]
RP   DOMAINS.
RA   Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA   Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA   Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT   "Sequence, structure, receptor-binding domains and internal repeats of
RT   human apolipoprotein B-100.";
RL   Nature 323:738-742(1986).
RN   [15]
RP   CALCIUM-BINDING DATA.
RX   MEDLINE=86242245; PubMed=3087360;
RA   Dashti N., Lee D.M., Mok T.;
RT   "Apolipoprotein B is a calcium binding protein.";
RL   Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN   [16]
RP   PALMITOYLATION OF CYS-1112.
RX   MEDLINE=20143590; PubMed=10679026;
RA   Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT   "Palmitoylation of apolipoprotein B is required for proper
RT   intracellular sorting and transport of cholesteroyl esters and
RT   triglycerides.";
RL   Mol. Biol. Cell 11:721-734(2000).
RN   [17]
RP   VARIANT SER-4338.
RX   MEDLINE=91071750; PubMed=1979313;
RA   Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA   Cunny G., Cambien F., Roizes G.;
RT   "Detection by denaturing gradient gel electrophoresis of a new
RT   polymorphism in the apolipoprotein B gene.";
RL   Hum. Genet. 86:91-93(1990).
RN   [18]
RP   VARIANT FDB GLN-3527.
RX   MEDLINE=89098975; PubMed=2563166;
RA   Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA   McCarthy B.J.;
RT   "Association between a specific apolipoprotein B mutation and familial
RT   defective apolipoprotein B-100.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN   [19]
RP   VARIANT LEU-2739.
RX   MEDLINE=91016974; PubMed=2216805;
RA   Huang L.-S., Gavish D., Breslow J.L.;
RT   "Sequence polymorphism in the human apoB gene at position 8344.";
RL   Nucleic Acids Res. 18:5922-5922(1990).
RN   [20]
RP   VARIANT FDB CYS-3558.
RX   MEDLINE=95190020; PubMed=7883971;
RA   Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA   Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT   "Familial ligand-defective apolipoprotein B. Identification of a new
RT   mutation that decreases LDL receptor binding affinity.";
RL   J. Clin. Invest. 95:1225-1234(1995).
RN   [21]
RP   VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP   AND THR-4481.
RX   MEDLINE=97044521; PubMed=8889592;
RA   Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA   Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT   "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT   PCR-SSCP.";
RL   Hum. Mutat. 8:282-285(1996).
RN   [22]
RP   VARIANTS FDB GLN-3527 AND CYS-3558.
RX   MEDLINE=97403938; PubMed=9259199;
RA   Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA   Krempf M., Giraudet P., Junien C., Boileau C.;
RT   "Familial ligand-defective apolipoprotein B-100: simultaneous
RT   detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a French
RT   population.";
RL   Hum. Mutat. 10:160-163(1997).
RN   [23]
RP   VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP   AND ILE-3921.
RX   MEDLINE=98141125; PubMed=9490296;
RA   Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT   "Screening for mutations of the apolipoprotein B gene causing
RT   hypocholesterolemia.";
RL   Hum. Genet. 102:44-49(1998).
CC   -!- FUNCTION: Apolipoprotein B is a major protein constituent of
CC       chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo
CC       B-100 functions as a recognition signal for the cellular binding
CC       and internalization of LDL particles by the apoB/E receptor.
CC   -!- SUBCELLULAR LOCATION: Secreted.

  Query Match             89.8%;  Score 44;  DB 1;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 35;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 9
Q7Z600
ID   Q7Z600      PRELIMINARY;      PRT;  4563 AA.
AC   Q7Z600;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Apolipoprotein B (Including Ag(X) antigen).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA   Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY324608; AAP72970.1; -.
DR   GO; GO:0005319; F:lipid transporter activity; IEA.
DR   GO; GO:0006869; P:lipid transport; IEA.
DR   InterPro; IPR009454; DUF1081.
DR   InterPro; IPR001747; Lipid_transprt_N.
DR   Pfam; PF06448; DUF1081; 1.
DR   Pfam; PF01347; Vitellogenin_N; 1.
```

```
DR   SMART; SM00638; LPD_N; 1.
KW   Lipoprotein.
SQ   SEQUENCE   4563 AA;  515553 MW;  030B34167CEDC63C CRC64;

  Query Match             89.8%;  Score 44;  DB 2;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 35;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 10
Q6PIM8
ID   Q6PIM8      PRELIMINARY;      PRT;   164 AA.
AC   Q6PIM8;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   ACSL3 protein (Fragment).
GN   Name=ACSL3;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC032144; AAH32144.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   164 AA;  18619 MW;  0BB00FE1A649E9AA CRC64;

  Query Match             83.7%;  Score 41;  DB 2;  Length 164;
  Best Local Similarity   80.0%;  Pred. No. 4.2;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db         77 TELARKKGLK 86


RESULT 11
AAH32144
ID   AAH32144    PRELIMINARY;      PRT;   164 AA.
AC   AAH32144;
DT   03-MAR-2004 (TrEMBLrel. 27, Created)
DT   03-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   03-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   ACSL3 protein (Fragment).
GN   ACSL3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC032144; AAH32144.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   164 AA;  18619 MW;  0BB00FE1A649E9AA CRC64;

  Query Match             83.7%;  Score 41;  DB 2;  Length 164;
  Best Local Similarity   80.0%;  Pred. No. 4.2;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db         77 TELARKKGLK 86


RESULT 12
LCF3_HUMAN
ID   LCF3_HUMAN      STANDARD;      PRT;   720 AA.
AC   O95573;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Long-chain-fatty-acid--CoA ligase 3 (EC 6.2.1.3) (Long-chain acyl-CoA
DE   synthetase 3) (LACS 3).
GN   Name=ACSL3; Synonyms=FACL3, ACS3, LACS3;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97321062; PubMed=9177793;
RA   Minekura H., Fujino T., Kang M.-J., Fujita T., Endo Y., Yamamoto T.T.;
RT   "Human acyl-coenzyme A synthetase 3 cDNA and localization of its gene
RT   (ACS3) to chromosome band 2q34-q35.";
RL   Genomics 42:180-181(1997).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21564184; PubMed=11707336; DOI=10.1016/S0378-1119(01)00714-4;
RA   Minekura H., Kang M.-J., Inagaki Y., Suzuki H., Sato H., Fujino T.,
RA   Yamamoto T.T.;
RT   "Genomic organization and transcription units of the human acyl-CoA
RT   synthetase 3 gene.";
RL   Gene 278:185-192(2001).
CC   -!- FUNCTION: Activation of long-chain fatty acids for both synthesis
CC       of cellular lipids, and degradation via beta-oxidation.
CC       Preferentially uses myristate, laurate, arachidonate and
CC       eicosapentaenoate as substrates (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + a long-chain carboxylic acid + CoA = AMP
CC       + diphosphate + an acyl-CoA.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SUBCELLULAR LOCATION: Microsomes, outer mitochondrial membrane and
CC       peroxisomal membrane.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D89053; BAA37142.1; -.
DR   EMBL; AB061712; BAB72074.1; -.
DR   Genew; HGNC:3570; ACSL3.
DR   MIM; 602371; -.
DR   GO; GO:0004321; F:fatty-acyl-CoA synthase activity; TAS.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Fatty acid metabolism; Ligase; Magnesium; Multigene family.
SQ   SEQUENCE   720 AA;  80345 MW;  845959A765BC6BF6 CRC64;

  Query Match             83.7%;  Score 41;  DB 1;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 20;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 13
Q8IUM9
ID   Q8IUM9      PRELIMINARY;      PRT;   720 AA.
AC   Q8IUM9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Acyl-CoA synthetase long-chain family member 3.
GN   Name=ACSL3;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Strausberg R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC       family.
DR   EMBL; BC041692; AAH41692.1; -.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
SQ   SEQUENCE   720 AA;  80419 MW;  AAC4B0B4543EC8DD CRC64;

  Query Match             83.7%;  Score 41;  DB 2;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 20;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 14
BAB72139
ID   BAB72139    PRELIMINARY;      PRT;   720 AA.
AC   BAB72139;
DT   02-MAR-2004 (TrEMBLrel. 27, Created)
DT   02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   Acyl-CoA synthetase 3.
GN   FACL3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Minekura H., Kang M.-J., Inagaki Y., Suzuki H., Sato H., Fujino T.,
RA   Yamamoto T.T.;
RT   "Genomic organization and transcription units of the human acyl-CoA
RT   synthetase 3 gene.";
RL   Gene 278:185-192(2001).
DR   EMBL; AB061436; BAB72139.1; -.
DR   EMBL; AB061423; BAB72139.1; JOINED.
DR   EMBL; AB061424; BAB72139.1; JOINED.
DR   EMBL; AB061425; BAB72139.1; JOINED.
DR   EMBL; AB061426; BAB72139.1; JOINED.
DR   EMBL; AB061427; BAB72139.1; JOINED.
DR   EMBL; AB061428; BAB72139.1; JOINED.
DR   EMBL; AB061429; BAB72139.1; JOINED.
DR   EMBL; AB061430; BAB72139.1; JOINED.
DR   EMBL; AB061431; BAB72139.1; JOINED.
DR   EMBL; AB061432; BAB72139.1; JOINED.
DR   EMBL; AB061433; BAB72139.1; JOINED.
DR   EMBL; AB061434; BAB72139.1; JOINED.
DR   EMBL; AB061435; BAB72139.1; JOINED.
SQ   SEQUENCE   720 AA;  80345 MW;  845959A765BC6BF6 CRC64;
```

```
  Query Match             83.7%;  Score 41;  DB 2;  Length 720;
  Best Local Similarity   80.0%;  Pred. No. 20;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              ||| ||:|||
Db        633 TELARKKGLK 642


RESULT 15
Q7YQQ0
ID   Q7YQQ0      PRELIMINARY;      PRT;   263 AA.
AC   Q7YQQ0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Procavia capensis (Cape hyrax) (Rock dassie).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Hyracoidea; Procaviidae; Procavia.
OX   NCBI_TaxID=9813;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548411; AAP97367.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     263    263
SQ   SEQUENCE   263 AA;  29532 MW;  536CF6149C1D062A CRC64;

  Query Match             81.6%;  Score 40;  DB 2;  Length 263;
  Best Local Similarity   80.0%;  Pred. No. 11;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TELTRKRGLK 10
              : ||||||||
Db        206 SSLTRKRGLK 215


Search completed: January 13, 2005, 01:51:02
Job time : 78.0328 secs
```

THIS PAGE IS BLANK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:14:40 ; Search time 77.2131 Seconds
(without alignments)
46.460 Million cell updates/sec

Title: US-09-823-418-12
Perfect score: 50
Sequence: 1 TDLTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	50	100.0	10	2	AAY30693	Aay30693 Apo-B100
2	46	92.0	10	2	AAY30692	Aay30692 Apo-B100
3	42	84.0	11	2	AAW57205	Aaw57205 Apo B bin
4	42	84.0	13	2	AAW57207	Aaw57207 Apo B 100
5	42	84.0	15	2	AAW41261	Aaw41261 Apolipopr
6	42	84.0	15	2	AAW96892	Aaw96892 ApoB-100
7	42	84.0	20	6	ABJ37575	Abj37575 Heparin b
8	42	84.0	22	2	AAW57208	Aaw57208 Apo B 100
9	42	84.0	22	2	AAW57209	Aaw57209 Apo B 100
10	42	84.0	34	5	AAE14541	Aae14541 Human apo
11	42	84.0	36	2	AAW96876	Aaw96876 Nucleic a
12	42	84.0	37	2	AAW64587	Aaw64587 Human apo
13	42	84.0	51	2	AAW96845	Aaw96845 Nucleic a
14	42	84.0	343	4	ABB37687	Abb37687 Peptide #
15	42	84.0	343	4	ABG52504	Abg52504 Human liv
16	42	84.0	377	2	AAR72704	Aar72704 Human apo
17	42	84.0	377	2	AAR34031	Aar34031 Sequence
18	42	84.0	2463	8	ADJ57400	Adj57400 Human apo
19	42	84.0	3923	2	AAY31237	Aay31237 Human Apo
20	42	84.0	4536	2	AAW41262	Aaw41262 Apolipopr
21	42	84.0	4536	2	AAW96826	Aaw96826 Amino aci
22	42	84.0	4560	5	AAU98981	Aau98981 Human apo
23	42	84.0	4561	7	ADD48677	Add48677 Human Pro
24	42	84.0	4563	5	AAO15893	Aao15893 Human apo
25	42	84.0	4563	6	ABR40253	Abr40253 Human ali
26	42	84.0	4563	6	ABU79140	Abu79140 Apolipopr
27	42	84.0	4563	7	ADF43408	Adf43408 Apolipopr
28	42	84.0	4563	8	ADH18871	Adh18871 Human apo
29	42	84.0	4563	8	ADH18870	Adh18870 Human apo
30	42	84.0	4563	8	ADO33445	Ado33445 Human apo
31	42	84.0	4563	8	ADO33447	Ado33447 Human apo
32	42	84.0	4590	4	AAU33184	Aau33184 Novel hum
33	40	80.0	782	7	ADA06278	Ada06278 Human cel
34	39	78.0	11	2	AAW57206	Aaw57206 Apo B 100
35	39	78.0	11	2	AAW87717	Aaw87717 Analogue
36	39	78.0	11	5	AAE21732	Aae21732 BSMR effe
37	39	78.0	11	6	ABU07938	Abu07938 Apoprotei
38	39	78.0	11	7	ADF56451	Adf56451 Human apo
39	39	78.0	12	2	AAW41260	Aaw41260 Apolipopr
40	39	78.0	15	2	AAW22911	Aaw22911 Low densi
41	39	78.0	23	6	ABR57177	Abr57177 Human PDG
42	38	76.0	10	2	AAY30682	Aay30682 Apo-B100
43	38	76.0	10	2	AAY30687	Aay30687 Apo-B100
44	38	76.0	63	4	ABG09607	Abg09607 Novel hum
45	38	76.0	465	4	AAB92994	Aab92994 Human pro

ALIGNMENTS

RESULT 1
AAY30693
ID AAY30693 standard; peptide; 10 AA.
XX
AC AAY30693;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can

CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 100.0%; Score 50; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
||||||||||
Db 1 TDLTRKRGLK 10

RESULT 2
AAY30692
ID AAY30692 standard; peptide; 10 AA.
XX
AC AAY30692;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 92.0%; Score 46; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.038;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
|:||||||||
Db 1 TELTRKRGLK 10

RESULT 3
AAW57205
ID AAW57205 standard; peptide; 11 AA.
XX
AC AAW57205;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B binding site peptide 2.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 12; Page 52; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B binding site
CC peptide which can be used as a component of a non-naturally occurring,
CC receptor-competent low density lipoprotein (LDL) particle of the present
CC invention. The LDL particle comprises at least 1 peptide component that
CC has at least 1 binding site for an apo B protein receptor and at least 1
CC lipophilic substituent. Also described in the invention are peptides
CC containing an apo B binding sequence with at least 70% identity with
CC sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their dimers. Non-
CC naturally occurring, receptor-competent LDL particles are useful as: (i)
CC drug-targeting vectors for delivering anticancer drugs to cancer cells
CC that express an apo B protein receptor, and (ii) additives for cell
CC culture media especially as growth supplements. Non-naturally occurring,
CC receptor-competent LDL particles do not require the complete apo B
CC sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 11 AA;

Query Match 84.0%; Score 42; DB 2; Length 11;
Best Local Similarity 90.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
| ||||||||
Db 2 TRLTRKRGLK 11

```
RESULT 4
AAW57207
ID   AAW57207 standard; peptide; 13 AA.
XX
AC   AAW57207;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide B.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 13 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 13;
  Best Local Similarity   90.0%;  Pred. No. 0.3;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db          3 TRLTRKRGLK 12


RESULT 5
AAW41261
ID   AAW41261 standard; peptide; 15 AA.
XX
AC   AAW41261;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Apolipoprotein B-100 fragment.
XX
KW   Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW   thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
KW   angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW   prothrombinase complex.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9743311-A1.
XX
PD   20-NOV-1997.
XX
PF   09-MAY-1997;   97WO-GB001255.
XX
PR   09-MAY-1996;   96GB-00009702.
XX
PA   (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI   Bruckdorfer KR,  Ettelaie C;
XX
DR   WPI; 1998-008798/01.
XX
PT   Peptide fragments of apo:lipo:protein B-100 with anticoagulant activity -
PT   used for treating or preventing coagulation, inhibiting angiogenesis,
PT   cell differentiation and apoptosis.
XX
PS   Disclosure; Page 22; 60pp; English.
XX
CC   This sequence is an example of the peptide of the invention. It has the
CC   formula (I), or their variants with one or more internal deletions,
CC   insertions or substitutions, while retaining anti-coagulant properties of
CC   apolipoprotein B-100 (apoB-100). Z1-KAQ-X1-KKNKHRHS-X2-T-Z2 (I) X1 = S or
CC   Y; X2 = T or I; Z1 = the N terminus of the peptide, or 1-47 amino acids
CC   (aa); Z2 = the C terminus of the peptide, a terminal amide group or 1-77
CC   aa. Compositions containing the peptide are used for simultaneous,
CC   separate or sequential treatment of cancer, particularly to prevent
CC   metastatic spread. They are also used to inhibit thromboplastin-mediated
CC   processes, specifically to prevent or reduce blood coagulation (e.g.
CC   during or after surgery or in cases of heart attack, stroke etc.) and to
CC   inhibit angiogenesis, cellular differentiation or apoptosis. KRAD-14,
CC   which is active as such or as part of a 98-aa peptide, inhibits
CC   activation of the prothrombinase complex; and prevents activation of
CC   factor VII on the surface of thromboplastin and of platelets by thrombin.
CC   It binds to the residues 58-66 of thromboplastin. Since (I) are much
CC   smaller than apoB-100, they act more quickly
XX
SQ   Sequence 15 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 15;
  Best Local Similarity   90.0%;  Pred. No. 0.34;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db          1 TRLTRKRGLK 10


RESULT 6
AAW96892
ID   AAW96892 standard; peptide; 15 AA.
XX
AC   AAW96892;
XX
DT   22-APR-1999  (first entry)
XX
DE   ApoB-100 nuclear localisation signal sequence, residues 3353-3367.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
```

KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Claim 19; Fig 13D; 293pp; English.
XX
CC AAW96878-97 represent nuclear localisation signal sequence derived from
CC human apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein
CC component of very-low density lipoproteins (VLDL), intermediate density
CC lipoprotein (IDL), low density lipoproteins ((LDL) and lipoprotein a. The
CC present sequence can be used in the composition of the invention. The
CC specification describes a composition that comprises LDL and
CC apolipoproteins for the binding and in vivo transport of nucleic acids.
CC The composition is used to deliver nucleic acids to eukaryotic cells, in
CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC molecule (or ribozyme). Specifically they are used for gene therapy of
CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC fibrosis and arteriosclerosis
XX
SQ Sequence 15 AA;

Query Match 84.0%; Score 42; DB 2; Length 15;
Best Local Similarity 90.0%; Pred. No. 0.34;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
| ||||||||
Db 6 TRLTRKRGLK 15

RESULT 7
ABJ37575
ID ABJ37575 standard; peptide; 20 AA.
XX
AC ABJ37575;
XX
DT 10-MAY-2003 (first entry)
XX
DE Heparin binding peptide sequence #28.
XX
KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW cardiovascular; circulatory; ligand; sulphated; sulphonated; tumour;
KW rheumatoid arthritis; hypoxia; diabetic retinopathy; heparin.
XX
OS Unidentified.
XX
PN WO2003007689-A2.
XX
PD 30-JAN-2003.
XX
PF 22-JUL-2002; 2002WO-US023419.
XX
PR 20-JUL-2001; 2001US-0306726P.
XX
PA (ETHZ-) ETH ZUERICH.
PA (UYZU-) UNIV ZURICH.
XX
PI Hubbell JA, Schoenmakers R, Maynard HD;
XX
DR WPI; 2003-300420/29.
XX
PT Use of a ligand comprising of at least one sulfated or sulfonated amino
PT acid for the treatment of e.g. tumors, rheumatoid arthritis, diabetic
PT retinopathy and hypoxia.
XX
PS Disclosure; Fig 2; 79pp; English.
XX
CC The invention relates to a novel ligand for binding a target biomolecule,
CC which comprises a peptide having at least one sulphated or sulphonated
CC amino acid and at least one amino acid chosen from neutral and positively
CC charged amino acids. The novel ligands can be used for the treatment of
CC e.g. tumours, rheumatoid arthritis, diabetic retinopathy and hypoxia.
CC This sequence represents a heparin binding peptide relating to the
CC invention
XX
SQ Sequence 20 AA;

Query Match 84.0%; Score 42; DB 6; Length 20;
Best Local Similarity 90.0%; Pred. No. 0.46;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
| ||||||||
Db 7 TRLTRKRGLK 16

RESULT 8
AAW57208
ID AAW57208 standard; peptide; 22 AA.
XX
AC AAW57208;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide C.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
FT Modified-site 22
FT /note= "attached to cholesterol"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.

XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 84.0%; Score 42; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 0.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
| ||||||||
Db 7 TRLTRKRGLK 16

RESULT 9
AAW57209
ID AAW57209 standard; peptide; 22 AA.
XX
AC AAW57209;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide D.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 84.0%; Score 42; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 0.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
| ||||||||
Db 7 TRLTRKRGLK 16

RESULT 10
AAE14541
ID AAE14541 standard; peptide; 34 AA.
XX
AC AAE14541;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human apoB-100 derived peptide p62.
XX
KW Oxidised low density lipoprotein; oxLDL; immunoassay; periaortitis;
KW cardiovascular disease; coronary heart disease; pre-eclampsia;
KW non-insulin-dependent diabetes; endothelial dysfunction; human; apoB-100;
KW peptide p62.
XX
OS Homo sapiens.
XX
PN WO200206314-A2.
XX
PD 24-JAN-2002.
XX
PF 18-JUL-2001; 2001WO-GB003212.
XX
PR 18-JUL-2000; 2000GB-00017641.
XX
PA (ARKT-) ARK THERAPEUTICS LTD.
XX
PI Narvanen O, Yla-Herttuala S;
XX
DR WPI; 2002-179777/23.
XX
PT New peptide useful in enzyme immunoassays for detecting oxidized low
PT density lipoprotein which is a marker of coronary heart disease and other
PT cardiovascular diseases, has affinity for oxidized low density
PT lipoprotein.
XX
PS Claim 6; Page 5; 21pp; English.
XX
CC The invention relates to peptides having affinity for oxidised low
CC density lipoprotein (oxLDL), in cyclised or multimeric form. The peptide
CC is useful in an immunoassay to determine the presence, and optionally,
CC the amount of antibodies in a sample, having affinity for oxLDL.
CC Preferably immobilised peptide is useful for measuring the amount of
CC autoantibodies for oxLDL in a sample, especially a serum or plasma sample
CC from a patient for evaluating the risk of coronary heart diseases, other
CC cardiovascular diseases, and several other disorders such as
CC periaortitis, pre-eclampsia, non-insulin-dependent diabetes and
CC endothelial dysfunction. The peptide of the invention is stable, can be
CC synthesised easily without the need to isolate proteins from a patient's

```
CC   blood, and has a long half-life. The present sequence is human apoB-100
CC   derived peptide p62 used in the invention
XX
SQ   Sequence 34 AA;

  Query Match             84.0%;  Score 42;  DB 5;  Length 34;
  Best Local Similarity   90.0%;  Pred. No. 0.78;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db         25 TRLTRKRGLK 34


RESULT 11
AAW96876
ID   AAW96876 standard; peptide; 36 AA.
XX
AC   AAW96876;
XX
DT   22-APR-1999  (first entry)
XX
DE   Nucleic acid binding domain from apoB-100, residues 3348-3390.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9856938-A1.
XX
PD   17-DEC-1998.
XX
PF   10-JUN-1998;   98WO-US011927.
XX
PR   13-JUN-1997;   97US-00874807.
PR   14-MAY-1998;   98US-00079030.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Guevara JG,  Hoogeveen RC,  Moore JP;
XX
DR   WPI; 1999-070331/06.
XX
PT   Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT   used for delivering nucleic acid to cells for gene therapy and antisense
PT   treatment.
XX
PS   Claim 16; Fig 12C; 293pp; English.
XX
CC   AAW96827-77 represent nucleic acid binding domains derived from human
CC   apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC   of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC   (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC   sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 36 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 36;
  Best Local Similarity   90.0%;  Pred. No. 0.82;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db         11 TRLTRKRGLK 20


RESULT 12
AAW64587
ID   AAW64587 standard; peptide; 37 AA.
XX
AC   AAW64587;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Human apolipoprotein peptide fragment #1.
XX
KW   Factor V; human; detection; protein function; blood coagulation; apo;
KW   fat metabolism; Leyden mutation; deep vein thrombosis; apolipoprotein;
KW   Alzheimer's disease; 5,10-methylenetetrahydrofolate reductase; prion;
KW   hypercysteinaemia; factor VII; cardiovascular disease; pathogen; virus.
XX
OS   Homo sapiens.
XX
PN   EP857973-A2.
XX
PD   12-AUG-1998.
XX
PF   12-JAN-1998;  98EP-00890007.
XX
PR   13-JAN-1997;  97AT-00000044.
XX
PA   (IMMO ) IMMUNO AG.
XX
PI   Moritz B,  Kiessig S,  Lang H,  Schenk V;
XX
DR   WPI; 1998-416142/36.
XX
PT   Detecting or quantifying mutant protein in presence of wild-type protein
PT   comprises reaction with ligand - used to detect mutant blood coagulation
PT   factors or apolipoproteins for diagnosing risk of thrombosis.
XX
PS   Example 2; Page 9; 18pp; German.
XX
CC   AAW64587 and AAW64588 are fragments of human apolipoprotein which are
CC   used with Factor V protein fragments in a novel method to detect the
CC   presence of a mutated protein in a sample that may also contain the
CC   corresponding wild-type protein. The method is used to detect mutations
CC   that alter protein functions (either point mutation or small insertions
CC   or deletions), particularly in proteins involved in blood coagulation or
CC   metabolism of fat. Protein functions which are specially detectable are
CC   the Leyden mutation in factor V (associated with increased risk of deep
CC   vein thrombosis), mutations in apolipoprotein (apo) genes (certain
CC   alleles of apoE indicates increased risk of developing Alzheimer's
CC   disease), thermostable 5,10-methylenetetrahydrofolate reductase
CC   (associated with hypercysteinaemia and venous thrombosis) and factor VII
CC   mutations (associated with increased risk of cardiovascular disease). The
CC   method can also be applied to proteins from pathogens, e.g. viruses or
CC   prions. The method does not require complex apparatus for polymerase
CC   chain reactions, it is simple, standardisable and reliable and is
CC   particularly suited to routine screening. It also allows mutant protein
CC   in a sample to be quantified
XX
SQ   Sequence 37 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 37;
  Best Local Similarity   90.0%;  Pred. No. 0.85;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db         11 TRLTRKRGLK 20


RESULT 13
AAW96845
ID   AAW96845 standard; peptide; 51 AA.
```

```
XX
AC   AAW96845;
XX
DT   22-APR-1999  (first entry)
XX
DE   Nucleic acid binding domain from apoB-100.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9856938-A1.
XX
PD   17-DEC-1998.
XX
PF   10-JUN-1998;   98WO-US011927.
XX
PR   13-JUN-1997;   97US-00874807.
PR   14-MAY-1998;   98US-00079030.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Guevara JG,  Hoogeveen RC,  Moore JP;
XX
DR   WPI; 1999-070331/06.
XX
PT   Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT   used for delivering nucleic acid to cells for gene therapy and antisense
PT   treatment.
XX
PS   Claim 16; Page 151; 293pp; English.
XX
CC   AAW96827-77 represent nucleic acid binding domains derived from human
CC   apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC   of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC   (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC   sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 51 AA;

  Query Match             84.0%;  Score 42;  DB 2;  Length 51;
  Best Local Similarity   90.0%;  Pred. No. 1.2;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db          6 TRLTRKRGLK 15


RESULT 14
ABB37687
ID   ABB37687 standard; peptide; 343 AA.
XX
AC   ABB37687;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #5193 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 27; SEQ ID NO 30322; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a peptide encoded by a single exon nucleic acid probe
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 343 AA;

  Query Match             84.0%;  Score 42;  DB 4;  Length 343;
  Best Local Similarity   90.0%;  Pred. No. 7.8;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        169 TRLTRKRGLK 178


RESULT 15
ABG52504
ID   ABG52504 standard; peptide; 343 AA.
XX
AC   ABG52504;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human liver peptide, SEQ ID No 31152.
XX
KW   Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW   hypercholesterolaemia; coronary heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
```

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 27; SEQ ID NO 31152; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 343 AA;

Query Match 84.0%; Score 42; DB 4; Length 343;
Best Local Similarity 90.0%; Pred. No. 7.8;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        169 TRLTRKRGLK 178
```

Search completed: January 13, 2005, 01:43:02
Job time : 78.3798 secs

THIS PAGE LEFT BLANK

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:30:05 ; Search time 14.4262 Seconds
(without alignments)
66.696 Million cell updates/sec

Title: US-09-823-418-12
Perfect score: 50
Sequence: 1 TDLTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	88.0	275	2	E60950	apolipoprotein B-1
2	42	84.0	596	2	S32802	apolipoprotein B -
3	42	84.0	4563	1	LPHUB	apolipoprotein B-1
4	40	80.0	274	2	A60950	apolipoprotein B-1
5	39	78.0	269	2	C60950	apolipoprotein B-1
6	39	78.0	779	2	JH0102	apolipoprotein B -
7	37	74.0	1091	2	T35822	probable regulator
8	36	72.0	266	2	S60674	hypothetical prote
9	35	70.0	187	2	T21671	hypothetical prote
10	35	70.0	309	1	E65112	hypothetical 34.6
11	35	70.0	309	2	E85985	hypothetical prote
12	35	70.0	309	2	B91140	hypothetical prote
13	35	70.0	382	1	A44056	nucleocapsid prote
14	35	70.0	1217	2	T25894	hypothetical prote
15	35	70.0	3450	2	T26963	hypothetical prote
16	35	70.0	3461	2	T26964	hypothetical prote
17	34	68.0	168	2	T03168	hypothetical prote
18	34	68.0	206	2	T01788	aminoglycoside 6'-
19	34	68.0	211	2	F75474	hypothetical prote
20	34	68.0	493	2	E71008	hypothetical prote
21	34	68.0	680	2	S29682	DNA topoisomerase
22	34	68.0	684	2	S29683	DNA gyrase B, novo
23	34	68.0	686	2	T10969	DNA topoisomerase
24	34	68.0	1058	2	S65460	apolipoprotein B -
25	34	68.0	1253	1	A44400	myosin heavy chain
26	34	68.0	1254	2	A54818	myosin-VI [similar
27	34	68.0	1265	2	A59299	unconventional myo
28	34	68.0	1615	2	JC6510	ras-responsive ele
29	34	68.0	1778	2	JT0382	apolipoprotein B -
30	34	68.0	2629	2	I46569	apolipoprotein B -
31	33	66.0	102	2	D75018	hypothetical prote
32	33	66.0	254	2	H95070	hypothetical prote
33	33	66.0	264	2	F97938	hypothetical prote
34	33	66.0	325	2	B72475	probable transcrip
35	33	66.0	339	2	S62596	ubiquinol-cytochro
36	33	66.0	391	2	S60672	replication protei
37	33	66.0	402	2	S55980	probable membrane
38	33	66.0	784	2	JH0101	apolipoprotein B-1
39	33	66.0	2100	2	T03223	probable polyketid
40	32	64.0	70	2	B64497	hypothetical prote
41	32	64.0	149	2	E72338	conserved hypothet
42	32	64.0	214	2	AE2934	transcription regu
43	32	64.0	219	2	B98348	probable transcrip
44	32	64.0	252	2	AE0876	conserved hypothet
45	32	64.0	281	2	F82832	pantoate-beta-alan

ALIGNMENTS

RESULT 1
E60950
apolipoprotein B-100 - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: E60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LI
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: E60950
A;Molecule type: mRNA
A;Residues: 1-275 <LAW>
A;Cross-references: UNIPROT:Q7LZ77
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotei

Query Match 88.0%; Score 44; DB 2; Length 275;
Best Local Similarity 90.0%; Pred. No. 0.46;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        221 TSLTRKRGLK 230
```

RESULT 2
S32802
apolipoprotein B - crab-eating macaque (fragment)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S32802
R;Pape, M.E.; Castle, C.K.; Murray, R.W.; Funk, G.M.; Hunt, C.E.; Marotti, K.R.; Melchi
Biochim. Biophys. Acta 1086, 326-334, 1991
A;Title: Apo B metabolism in the cynomolgus monkey: evidence for post-transcriptional r
A;Reference number: S32802; MUID:92075708; PMID:1742325
A;Accession: S32802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-596 <PAP>
A;Cross-references: UNIPROT:Q28473; EMBL:X15737; NID:g38047; PIDN:CAA33755.1; PID:g9301
C;Superfamily: apolipoprotein B

Query Match 84.0%; Score 42; DB 2; Length 596;
Best Local Similarity 90.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        226 TRLTRKRGLK 235
```

RESULT 3
LPHUB
apolipoprotein B-100 precursor - human
N;Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
C;Species: Homo sapiens (man)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: A27850; A25679; A25263; A25267; A25266; A24320; A24684; A23817; A25774; A26
4452; I61909; I59510; I39474; I39469; I84624; I37179; PS0058
R;Ludwig, E.H.; Blackhart, B.D.; Pierotti, V.R.; Caiati, L.; Fortier, C.; Knott, T.; Scc
DNA 6, 363-372, 1987
A;Title: DNA sequence of the human apolipoprotein B gene.
A;Reference number: A27850; MUID:88003974; PMID:3652907
A;Accession: A27850
A;Molecule type: DNA
A;Residues: 1-617,'A',619-1929,'F',1931-3318,'D',3320-3426,'T',3428-3431,'Q',3433-3731,'
A;Cross-references: UNIPROT:P04114; UNIPROT:P78482; UNIPROT:P78479; UNIPROT:Q9UMN0; UNIP
R;Cladaras, C.; Hadzopoulou-Cladaras, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A;Title: The complete sequence and structural analysis of human apolipoprotein B-100: re
A;Reference number: A91058; MUID:87161758; PMID:3030729
A;Accession: A25679
A;Molecule type: mRNA
A;Residues: 1-11,15-2539,'S',2541-3823,'R',3825-4563 <CLA>
A;Note: 1109-Asp was also found
R;Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McCa
Nucleic Acids Res. 14, 7501-7503, 1986
A;Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A;Reference number: A93639; MUID:87016385; PMID:3763409
A;Accession: A25263
A;Molecule type: mRNA
A;Residues: 1-272,'N',274-617,'A',619-1217,'E',1219-2091,'V',2093-2364,'T',2366-2679,'Q'
A;Cross-references: GB:X04506; NID:g34330; PIDN:CAA28191.1; PID:g34331
R;Law, S.W.; Grant, S.M.; Higuchi, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer Jr
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A;Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A;Reference number: A94134; MUID:87041416; PMID:3464946
A;Accession: A25267
A;Molecule type: mRNA
A;Residues: 1-617,'A',619-703,'P',705-792,'R',794-1270,'S',1272-1866,'G',1868-2036,'N',2
4189-4220,'M',4222-4563 <LAW>
A;Note: the codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4188, and
R;Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Gotto Jr., A.M
J. Biol. Chem. 261, 12918-12921, 1986
A;Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A;Reference number: A92556; MUID:87008488; PMID:3759943
A;Accession: A25266
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-328,'V',330-644,'I',646-918,'P',920-3318,'D',3320-3426,'T',3428-
9-4132,'G',4134-4180,'E',4182-4563 <CHE>
A;Cross-references: GB:J02610; NID:g178803; PIDN:AAA35549.1; PID:g178804
A;Note: a total of 2366 residues were confirmed by direct sequencing of tryptic peptides
R;Protter, A.A.; Hardman, D.A.; Sato, K.Y.; Schilling, J.W.; Yamanaka, M.; Hort, Y.J.; H
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A;Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein
A;Reference number: A24320; MUID:86287319; PMID:3461454
A;Accession: A24320
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-617,'A',619-941,'YYIWSLPPKP',951-1138,'PTGRLPNCFSNGLICYSLWLHSFQE
A;Cross-references: GB:M14081; NID:g178795; PIDN:AAA51752.1; PID:g553189
R;Law, S.W.; Lackner, K.J.; Hospattankar, A.V.; Anchors, J.M.; Sakaguchi, A.Y.; Naylor,
Proc. Natl. Acad. Sci. U.S.A. 82, 8340-8344, 1985
A;Title: Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of
A;Reference number: A24684; MUID:86094221; PMID:3001697
A;Accession: A24684
A;Molecule type: mRNA
A;Residues: 485-617,'A',619-1044 <LA2>
A;Cross-references: GB:M12480; NID:g178791; PIDN:AAA51751.1; PID:g178792
R;Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; Ki
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A;Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipopr
A;Reference number: A94088; MUID:86149325; PMID:3513177
A;Accession: A23817
A;Molecule type: mRNA
A;Residues: 1-291 <PRO>
A;Cross-references: GB:M12681; NID:g178797; PIDN:AAA51753.1; PID:g178798
R;Deeb, S.S.; Motulsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A;Title: A partial cDNA clone for human apolipoprotein B.
A;Reference number: A25774; MUID:85270450; PMID:3860836
A;Accession: A25774
A;Molecule type: mRNA
A;Residues: 709-791,'SSSWKAASHGCPHSAGD',810-906 <DEE>
A;Cross-references: GB:K03175; NID:g178821; PIDN:AAA51759.1; PID:g178822
R;Carlsson, P.; Darnfors, C.; Olofsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A;Title: Analysis of the human apolipoprotein B gene; complete structure of the B-74 re
A;Reference number: A91565; MUID:87191999; PMID:2883086
A;Accession: A26533
A;Molecule type: mRNA
A;Residues: 1282-2721,2742-3290,'L',3292-3336,'N',3338-3948,'F',3950-3963,'Y',3965-4180,
A;Cross-references: GB:M15421; NID:g178817; PIDN:AAA51758.1; PID:g178818
R;Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Sato, K.Y.; Lau, K.; Yaman
Biochemistry 26, 5478-5486, 1987
A;Title: Structural comparison of human apolipoproteins B-48 and B-100.
A;Reference number: A29671; MUID:88050832; PMID:3676265
A;Accession: A29671
A;Molecule type: mRNA
A;Residues: 1671-2323,'PYW',2327-2352,'H',2354-2398 <HAR>
A;Cross-references: GB:M17367; NID:g178731; PIDN:AAA51741.1; PID:g178732
R;Shoulders, C.C.; Myant, N.B.; Sidoli, A.; Rodriguez, J.C.; Cortese, C.; Baralle, F.E.
Atherosclerosis 58, 277-289, 1985
A;Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for more than o
A;Reference number: A90084; MUID:86130855; PMID:3841481
A;Accession: A29287
A;Molecule type: mRNA
A;Residues: 3846-4298 <SHO>
R;Pfitzner, R.; Wagener, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986
A;Title: Isolation, expression and characterization of a human apolipoprotein B 100-spe
A;Reference number: A25572; MUID:87076044; PMID:3024665
A;Accession: A25572
A;Molecule type: mRNA
A;Residues: 4219-4337,'S',4339-4563 <PFI>
A;Cross-references: GB:M36676
R;Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow, J.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269, 1985
A;Reference number: A24738; MUID:86042646; PMID:2932736
A;Accession: A24738
A;Molecule type: mRNA
A;Residues: 'N',3729-3731,'I',3733-3875,'A',3877-3948,'F',3950-3963,'Y',3965-3982,'S',3
A;Cross-references: GB:M12413; NID:g178735; PIDN:AAA51742.1; PID:g178736
R;Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.; Silberman, S.R.; Ca
Science 238, 363-366, 1987
A;Title: Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific i
A;Reference number: A40133; MUID:88018019; PMID:3659919
A;Accession: B40133
A;Molecule type: mRNA
A;Residues: 2165-2179 <CH1>
A;Cross-references: GB:M18036; NID:g178799; PIDN:AAA51754.1; PID:g178800
A;Note: this mRNA includes the stop codon of the organ-specific mRNA for apo48
A;Accession: A40133
A;Molecule type: protein
A;Residues: 51-75;101-110;129-139;158-174;197-207;276-287;298-304;306-314;526-532;538-5
36;1486-1498;1537-1556;1563-1572;1601-1610;1647-1661;1697-1724;1770-1781;1859-1897;1968
A;Note: these fragments were derived from apo48
R;Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987
A;Title: Carboxyl terminal analysis of human B-48 protein confirms the novel mechanism
A;Reference number: A28002; MUID:88106542; PMID:3426612
A;Accession: A28002
A;Molecule type: mRNA
A;Residues: 2129-2179,2181-2235 <HA2>
A;Cross-references: GB:M18471
A;Experimental source: intestine
A;Note: this mRNA from intestine includes a stop codon created by RNA editing in place
R;Mehrabian, M.; Schumaker, V.N.; Fareed, G.C.; West, R.; Johnson, D.F.; Kirchgessner,

Nucleic Acids Res. 13, 6937-6953, 1985
A;Title: Human apolipoprotein B: identification of cDNA clones and characterization of m
A;Reference number: A24269; MUID:86041888; PMID:3903660
A;Accession: A24269
A;Molecule type: mRNA
A;Residues: 3056-3159 <MEH>
A;Cross-references: GB:X03045; NID:g28783; PIDN:CAA26850.1; PID:g929609
R;Hospattankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer Jr., H.B.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A;Title: Identification of a novel in-frame translational stop codon in human intestine
A;Reference number: A29659; MUID:88049670; PMID:2445342
A;Accession: A29659
A;Molecule type: mRNA
A;Residues: 2169-2179 <HOS>
A;Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A;Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human intest
ch encodes the 250K apoB-48, CAA encoding 2180-Gln is substituted by the stop codon TAA,
R;Yang, C.; Kim, T.W.; Weng, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
A;Title: Isolation and characterization of sulfhydryl and disulfide peptides of human ap
A;Reference number: A35783; MUID:90319144; PMID:2115173
A;Contents: disulfide bonds
A;Accession: A35783
A;Molecule type: protein
A;Residues: 28-41;76-97,'I',99-100;175-193;206-215;239-249;259-266;357-399;455-490;512-5

A;Note: cysteines at positions 1112, 1422, 1505, 1662, 3761, 3917, and 4217 have free su
R;LeBoeuf, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
FEBS Lett. 170, 105-108, 1984
A;Title: Human apolipoprotein B: partial amino acid sequence.
A;Reference number: A22006; MUID:84208786; PMID:6373369
A;Accession: A22006
A;Molecule type: protein
A;Residues: 873-892,'K',894-896 <LE1>
A;Accession: B22006
A;Molecule type: protein
A;Residues: 3113,'L',3115-3130,'R',3132-3133,'P',3135-3136,'R' <LE2>
R;Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Caiati, L.; Onasch, M.A.; Wallis, S.C.;
J. Biol. Chem. 261, 15364-15367, 1986
A;Title: Structure of the human apolipoprotein B gene.
A;Reference number: A92564; MUID:87057153; PMID:2946672
A;Contents: annotation; gene structure
R;Wagener, R.; Pfitzner, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
A;Title: Studies on the organization of the human apolipoprotein B 100 gene.
A;Reference number: A90715; MUID:87271140; PMID:2886136
A;Contents: annotation; gene structure
R;Weisgraber, K.H.; Rall Jr., S.C.
J. Biol. Chem. 262, 11097-11103, 1987
A;Title: Human apolipoprotein B-100 heparin-binding sites.
A;Reference number: A92605; MUID:87280197; PMID:3301850
A;Contents: annotation; heparin binding and disulfide bond
R;Dashti, N.; Lee, D.M.; Mok, T.
Biochem. Biophys. Res. Commun. 137, 493-499, 1986
A;Title: Apolipoprotein B is a calcium binding protein.
A;Reference number: A90125; MUID:86242245; PMID:3087360
A;Contents: annotation; calcium binding
R;Carlsson, P.; Olofsson, S.O.; Bondjers, G.; Darnfors, C.; Wiklund, O.; Bjursell, G.
Nucleic Acids Res. 13, 8813-8826, 1985
A;Title: Molecular cloning of human apolipoprotein B cDNA.
A;Reference number: I37178; MUID:86093680; PMID:3841204
A;Accession: I37180

  Query Match 84.0%; Score 42; DB 1; Length 4563;
  Best Local Similarity 90.0%; Pred. No. 16;
  Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
  | ||||||||
Db 3385 TRLTRKRGLK 3394

RESULT 4
A60950
apolipoprotein B-100 - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 31-Dec-1993 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004
C;Accession: A60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LI
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: A60950
A;Molecule type: mRNA
A;Residues: 1-274 <LAW>
A;Cross-references: UNIPROT:Q7M2U9
A;Note: authors translated the codon GAT for residue 155 as His
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotei

  Query Match 80.0%; Score 40; DB 2; Length 274;
  Best Local Similarity 80.0%; Pred. No. 2.9;
  Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
  : ||||||||
Db 221 SSLTRKRGLK 230

RESULT 5
C60950
apolipoprotein B-100 - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: C60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LI
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: C60950
A;Molecule type: DNA
A;Residues: 1-269 <LAW>
A;Cross-references: UNIPROT:Q60537; UNIPROT:Q60536
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein

  Query Match 78.0%; Score 39; DB 2; Length 269;
  Best Local Similarity 100.0%; Pred. No. 4.5;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LTRKRGLK 10
  ||||||||
Db 218 LTRKRGLK 225

RESULT 6
JH0102
apolipoprotein B - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: JH0102
R;Smith, T.J.
submitted to GenBank, June 1990
A;Reference number: A38864
A;Accession: JH0102
A;Molecule type: DNA
A;Residues: 1-779 <SMI>
A;Cross-references: UNIPROT:Q60536; GB:M35187
A;Note: this is a revision to the sequence from reference JH0101
R;Smith, T.J.; Hautamaa, D.; Maeda, N.
Gene 87, 309-310, 1990
A;Title: Sequence of the putative low-density lipoprotein receptor-binding regions of a
A;Reference number: JH0101; MUID:90236327; PMID:2332175
A;Contents: annotation
A;Note: this sequence has been revised in reference A38864
C;Genetics:

A;Gene: apoB
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;
F;435-445/Region: receptor binding
F;646-656/Region: receptor binding

Query Match 78.0%; Score 39; DB 2; Length 779;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 LTRKRGLK 10
              ||||||||
Db        644 LTRKRGLK 651
```

RESULT 7
T35822
probable regulatory protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C;Accession: T35822
R;Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999
A;Reference number: Z21589
A;Accession: T35822
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1091 <MUR>
A;Cross-references: UNIPROT:Q9Z573; EMBL:AL035569; PIDN:CAB37582.1; GSPDB:GN00070; SCOED
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC8D9.18

Query Match 74.0%; Score 37; DB 2; Length 1091;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 DLTRKRGL 9
              ||||:|||
Db        729 DLTRRRGL 736
```

RESULT 8
S60674
hypothetical protein B - Corynebacterium glutamicum plasmid pGA1
C;Species: Corynebacterium glutamicum
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S60674
R;Nesvera, J.; Patek, M.; Hochmannova, J.; Abrhamova, Z.
submitted to the EMBL Data Library, August 1995
A;Description: Complete nucleotide sequence of the cryptic plasmid pGA1 from Corynebacte
A;Reference number: S60673
A;Accession: S60674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <NES>
A;Cross-references: UNIPROT:Q46059; EMBL:X90817; NID:g951006; PIDN:CAA62329.1; PID:g9510
C;Genetics:
A;Genome: plasmid pGA1

Query Match 72.0%; Score 36; DB 2; Length 266;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 DLTRKRGLK 10
              ||:||:|:|
Db        211 DLSRKKGIK 219
```

RESULT 9
T21671
hypothetical protein F32H2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21671
R;Kershaw, J.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19457
A;Accession: T21671
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-187 <WIL>
A;Cross-references: UNIPROT:P91866; EMBL:Z81523; PIDN:CAB04239.1; GSPDB:GN00019; CESP:F
A;Experimental source: clone F32H2
C;Genetics:
A;Gene: CESP:F32H2.6
A;Map position: 1
A;Introns: 22/2; 103/1

Query Match 70.0%; Score 35; DB 2; Length 187;
Best Local Similarity 70.0%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 TDLTRKRGLK 10
              ||| :||| |
Db         54 TDLPKKRGKK 63
```

RESULT 10
E65112
hypothetical 34.6 kD protein in arcB-gltB intergenic region - Escherichia coli (strain
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: E65112
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: E65112
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-309 <BLAT>
A;Cross-references: UNIPROT:P45476; GB:AE000400; GB:U00096; NID:g2367203; PIDN:AAC76243
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yhcC
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0486

Query Match 70.0%; Score 35; DB 1; Length 309;
Best Local Similarity 70.0%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 TDLTRKRGLK 10
              | | |:||||
Db        170 TQLARQRGLK 179
```

RESULT 11
E85985
hypothetical protein yhcC [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: E85985
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85985
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <STO>
A;Cross-references: UNIPROT:P45476; GB:AE005174; NID:g12517832; PIDN:AAG58345.1; GSPDB:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:

A;Gene: yhcC
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0486

  Query Match 70.0%; Score 35; DB 2; Length 309;
  Best Local Similarity 70.0%; Pred. No. 32;
  Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
Db 170 TQLARQRGLK 179

RESULT 12
B91140
hypothetical protein ECs4090 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: B91140
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B91140
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <HAY>
A;Cross-references: UNIPROT:P45476; GB:BA000007; PIDN:BAB37513.1; PID:g13363563; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4090
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0486

  Query Match 70.0%; Score 35; DB 2; Length 309;
  Best Local Similarity 70.0%; Pred. No. 32;
  Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TDLTRKRGLK 10
Db 170 TQLARQRGLK 179

RESULT 13
A44056
nucleocapsid protein - canine coronavirus (strain K378)
N;Alternate names: N protein
C;Species: canine coronavirus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: A44056
R;Vennema, H.; Rossen, J.W.A.; Wesseling, J.; Horzinek, M.C.; Rottier, P.J.M.
Virology 191, 134-140, 1992
A;Title: Genomic organization and expression of the 3' end of the canine and feline ente
A;Reference number: A44056; MUID:93033103; PMID:1329312
A;Accession: A44056
A;Molecule type: genomic RNA
A;Residues: 1-382 <VEN>
A;Cross-references: UNIPROT:Q04700; GB:X66717; NID:g58849; PIDN:CAA47246.1; PID:g58850
C;Genetics:
A;Gene: N
C;Superfamily: coronavirus nucleocapsid protein
C;Keywords: glycoprotein; nucleocapsid
F;28,134,154,172,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match 70.0%; Score 35; DB 1; Length 382;
  Best Local Similarity 75.0%; Pred. No. 39;
  Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DLTRKRGL 9
Db 13 DITKKRGL 20

RESULT 14
T25894
hypothetical protein T19B4.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T25894
R;Gattung, S.
submitted to the EMBL Data Library, November 1996
A;Description: The sequence of C. elegans cosmid T19B4.
A;Reference number: Z20106
A;Accession: T25894
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1217 <GAT>
A;Cross-references: UNIPROT:P91457; EMBL:U80438; PIDN:AAB37636.1; GSPDB:GN00019; CESP:
A;Experimental source: strain Bristol N2; clone T19B4
C;Genetics:
A;Gene: CESP:T19B4.2
A;Map position: 1
A;Introns: 66/1; 119/3; 321/3; 552/3; 1123/2; 1197/2

  Query Match 70.0%; Score 35; DB 2; Length 1217;
  Best Local Similarity 100.0%; Pred. No. 1.2e+02;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TDLTRKR 7
Db 290 TDLTRKR 296

RESULT 15
T26963
hypothetical protein ZK1151.2a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T26963; T27704
R;Harris, B.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z20292
A;Accession: T26963
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3450 <WIL>
A;Cross-references: EMBL:AL031637; PIDN:CAA21049.1; GSPDB:GN00019; CESP:ZK1151.2a
A;Experimental source: clone Y47H9B
R;Harris, B.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z20408
A;Accession: T27704
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3450 <WI2>
A;Cross-references: EMBL:Z93398; PIDN:CAB07725.1; GSPDB:GN00019; CESP:ZK1151.2a
A;Experimental source: clone ZK1151
C;Genetics:
A;Gene: CESP:ZK1151.2a
A;Map position: 1
A;Introns: 270/3; 757/2; 1105/1; 1259/3; 2312/2; 2613/1; 2825/2; 3180/1; 3217/3; 3257/3

  Query Match 70.0%; Score 35; DB 2; Length 3450;
  Best Local Similarity 75.0%; Pred. No. 3e+02;
  Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DLTRKRGL 9
Db 2762 DVTRKRGI 2769

Search completed: January 13, 2005, 01:52:39
Job time : 15.4262 secs

THIS PAGE IS BLANK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:15:45 ; Search time 78.0328 Seconds
(without alignments)
73.735 Million cell updates/sec

Title: US-09-823-418-12
Perfect score: 50
Sequence: 1 TDLTRKRGLK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	88.0	275	2	Q7LZ77	Q7lz77 gallus gall
2	44	88.0	387	2	Q7YQN2	Q7yqn2 phalanger o
3	44	88.0	400	2	Q7YQM9	Q7yqm9 ornithorhyn
4	44	88.0	405	2	Q7YQN0	Q7yqn0 tachyglossu
5	42	84.0	414	2	Q7YQR5	Q7yqr5 aotus vocif
6	42	84.0	596	2	Q28473	Q28473 macaca fasc
7	42	84.0	3262	2	Q13788	Q13788 homo sapien
8	42	84.0	4563	1	APB_HUMAN	P04114 homo sapien
9	42	84.0	4563	2	Q7Z600	Q7z600 homo sapien
10	40	80.0	263	2	Q7YQQ0	Q7yqq0 procavia ca
11	40	80.0	274	2	Q7M2U9	Q7m2u9 oryctolagus
12	40	80.0	304	2	Q7YQP9	Q7yqp9 echinops te
13	40	80.0	314	2	Q7YQN8	Q7yqn8 ictonyx str
14	40	80.0	316	2	Q7YQP3	Q7yqp3 nandinia bi
15	40	80.0	318	2	Q7YQN9	Q7yqn9 zalophus ca
16	40	80.0	319	2	Q7YQP0	Q7yqp0 vulpes vulp
17	40	80.0	319	2	Q7YQP2	Q7yqp2 panthera le
18	40	80.0	320	2	Q7YQP4	Q7yqp4 manis sp. k
19	40	80.0	322	2	Q7YQP5	Q7yqp5 manis sp. k
20	40	80.0	339	2	Q7YR05	Q7yr05 macroscelid
21	40	80.0	361	2	Q7YQP8	Q7yqp8 amblysomus
22	40	80.0	364	2	Q7YQQ1	Q7yqq1 dugong dugo
23	40	80.0	386	2	Q7YQR1	Q7yqr1 tupaia tana
24	40	80.0	392	2	Q7YR11	Q7yr11 tarsius syr
25	40	80.0	411	2	Q7YQP7	Q7yqp7 ochotona pr
26	40	80.0	422	2	Q7YR12	Q7yr12 talpa europ
27	40	80.0	423	2	Q7YQQ9	Q7yqq9 sorex monti
28	40	80.0	426	2	Q7YQR2	Q7yqr2 alces alces
29	40	80.0	429	2	Q7YQQ8	Q7yqq8 crocidura f
30	40	80.0	438	2	Q7YQR4	Q7yqr4 balaena mys
31	40	80.0	441	2	Q7YQR3	Q7yqr3 phocoenoide
32	40	80.0	443	2	Q7YQN5	Q7yqn5 megaderma l
33	40	80.0	443	2	Q7YQP6	Q7yqp6 lepus ameri
34	40	80.0	445	2	Q7YQN6	Q7yqn6 bradypus tr
35	40	80.0	445	2	Q7YQQ7	Q7yqq7 tapirus bai
36	40	80.0	445	2	Q7YQR0	Q7yqr0 cynocephalu
37	40	80.0	445	2	Q7YR07	Q7yr07 lemur catta
38	40	80.0	445	2	Q7YR14	Q7yr14 rhinolophus
39	39	78.0	421	2	Q7TN68	Q7tn68 glaucomys v
40	39	78.0	432	2	Q7YR10	Q7yr10 diceros bic
41	39	78.0	436	2	Q7YQM8	Q7yqm8 nyctimene a
42	39	78.0	438	2	Q7YQM7	Q7yqm7 pteropus hy
43	39	78.0	438	2	Q7YR04	Q7yr04 rousettus a
44	39	78.0	440	2	Q7YQN4	Q7yqn4 myotis veli
45	39	78.0	445	2	Q7YR08	Q7yr08 chaetophrac

ALIGNMENTS

RESULT 1
Q7LZ77
ID Q7LZ77 PRELIMINARY; PRT; 275 AA.
AC Q7LZ77;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Apolipoprotein B-100 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90324804; PubMed=2373961;
RA Law A., Scott J.;
RT "A cross-species comparison of the apolipoprotein B domain that binds
RT to the LDL receptor.";
RL J. Lipid Res. 31:1109-1120(1990).
DR PIR; E60950; E60950.
FT NON_TER 1 1
FT NON_TER 275 275
SQ SEQUENCE 275 AA; 30678 MW; B7D8DA054E04B255 CRC64;

Query Match 88.0%; Score 44; DB 2; Length 275;
Best Local Similarity 90.0%; Pred. No. 1.8;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        221 TSLTRKRGLK 230
```

RESULT 2
Q7YQN2
ID Q7YQN2 PRELIMINARY; PRT; 387 AA.
AC Q7YQN2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Apolipoprotein B 100 (Fragment).
GN Name=apoB-100;
OS Phalanger orientalis (gray cuscus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Phalanger.
OX NCBI_TaxID=42473;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22761261; PubMed=12878460;
RA Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT "A new phylogenetic marker, apolipoprotein B, provides compelling
RT evidence for eutherian relationships.";
RL Mol. Phylogenet. Evol. 28:225-240(2003).

```
DR   EMBL; AF548431; AAP97387.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     387    387
SQ   SEQUENCE   387 AA;  43230 MW;  8300A9D7C54B42B0 CRC64;

  Query Match             88.0%;  Score 44;  DB 2;  Length 387;
  Best Local Similarity   90.0%;  Pred. No. 2.5;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        260 TSLTRKRGLK 269


RESULT 3
Q7YQM9
ID   Q7YQM9      PRELIMINARY;      PRT;   400 AA.
AC   Q7YQM9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Ornithorhynchus anatinus (Duckbill platypus).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
OX   NCBI_TaxID=9258;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548434; AAP97390.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     400    400
SQ   SEQUENCE   400 AA;  44611 MW;  DC79873CA6D01CFA CRC64;

  Query Match             88.0%;  Score 44;  DB 2;  Length 400;
  Best Local Similarity   90.0%;  Pred. No. 2.6;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        260 TSLTRKRGLK 269


RESULT 4
Q7YQN0
ID   Q7YQN0      PRELIMINARY;      PRT;   405 AA.
AC   Q7YQN0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Tachyglossus aculeatus (Australian echidna).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Monotremata; Tachyglossidae; Tachyglossus.
OX   NCBI_TaxID=9261;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548433; AAP97389.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     405    405
SQ   SEQUENCE   405 AA;  44975 MW;  551A98557E8B081D CRC64;

  Query Match             88.0%;  Score 44;  DB 2;  Length 405;
  Best Local Similarity   90.0%;  Pred. No. 2.7;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        260 TSLTRKRGLK 269


RESULT 5
Q7YQR5
ID   Q7YQR5      PRELIMINARY;      PRT;   414 AA.
AC   Q7YQR5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Aotus vociferans (Spix's owl monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX   NCBI_TaxID=57176;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548396; AAP97352.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     414    414
SQ   SEQUENCE   414 AA;  45955 MW;  EEFA8492157E1BDE CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 414;
  Best Local Similarity   90.0%;  Pred. No. 7;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        258 TRLTRKRGLK 267


RESULT 6
Q28473
ID   Q28473      PRELIMINARY;      PRT;   596 AA.
AC   Q28473;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Apolipoprotein B (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92075708; PubMed=1742325;
RA   Pape M.E., Castle C.K., Murray R.W., Funk G.M., Hunt C.E.,
RA   Marotti K.R., Melchior G.W.;
RT   "Apo B metabolism in the cynomolgus monkey: evidence for post-
RT   transcriptional regulation.";
RL   Biochim. Biophys. Acta 1086:326-334(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Murray R.;
```

```
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X15737; CAA33755.1; -.
DR   PIR; S32802; S32802.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     596    596
SQ   SEQUENCE   596 AA;  66757 MW;  B13BBA74E25C3120 CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 596;
  Best Local Similarity   90.0%;  Pred. No. 10;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db        226 TRLTRKRGLK 235


RESULT 7
Q13788
ID   Q13788      PRELIMINARY;      PRT;  3262 AA.
AC   Q13788;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   APOB protein (Fragment).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87191999; PubMed=2883086;
RA   Carlsson P., Darnfors C., Olofsson S.O., Bjursell G.;
RT   "Analysis of the human apolipoprotein B gene; complete structure of
RT   the B-74 region.";
RL   Gene 49:29-51(1986).
DR   EMBL; M15421; AAA51758.1; -.
DR   PIR; A27850; LPHUB.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0005319; F:lipid transporter activity; NAS.
DR   GO; GO:0006869; P:lipid transport; NAS.
FT   NON_TER       1      1
SQ   SEQUENCE   3262 AA;  370140 MW;  56603BC0618DD40D CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 3262;
  Best Local Similarity   90.0%;  Pred. No. 63;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db       2084 TRLTRKRGLK 2093


RESULT 8
APB_HUMAN
ID   APB_HUMAN      STANDARD;      PRT;  4563 AA.
AC   P04114; O00502; Q13787;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
DE   B-48 (Apo B-48)].
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87016385; PubMed=3763409;
RA   Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
RA   Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
RT   "Complete cDNA and derived protein sequence of human apolipoprotein B-
RT   100.";
RL   Nucleic Acids Res. 14:7501-7503(1986).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT GLU-4181.
RX   MEDLINE=88003974; PubMed=3652907;
RA   Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
RA   Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
RT   "DNA sequence of the human apolipoprotein B gene.";
RL   DNA 6:363-372(1987).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANTS ILE-98 AND GLU-4181.
RX   MEDLINE=87008488; PubMed=3759943;
RA   Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
RA   Gotto A.M. Jr., Chan L.;
RT   "The complete cDNA and amino acid sequence of human apolipoprotein B-
RT   100.";
RL   J. Biol. Chem. 261:12918-12921(1986).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87041416; PubMed=3464946;
RA   Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
RA   Lee N., Brewer H.B. Jr.;
RT   "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
RT   derived amino acid sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87161758; PubMed=3030729;
RA   Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
RA   Zannis V.I.;
RT   "The complete sequence and structural analysis of human apolipoprotein
RT   B-100: relationship between apoB-100 and apoB-48 forms.";
RL   EMBO J. 5:3495-3507(1986).
RN   [6]
RP   SEQUENCE OF 709-906 FROM N.A.
RX   MEDLINE=85270450; PubMed=3860836;
RA   Deeb S.S., Motulsky A.G., Albers J.J.;
RT   "A partial cDNA clone for human apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN   [7]
RP   SEQUENCE OF 3056-3159 FROM N.A.
RX   MEDLINE=86041888; PubMed=3903660;
RA   Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
RA   Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
RT   "Human apolipoprotein B: identification of cDNA clones and
RT   characterization of mRNA.";
RL   Nucleic Acids Res. 13:6937-6953(1985).
RN   [8]
RP   SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A., AND VARIANT GLU-4181.
RX   MEDLINE=86093680; PubMed=3841204;
RA   Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
RA   Bjursell G.;
RT   "Molecular cloning of human apolipoprotein B cDNA.";
RL   Nucleic Acids Res. 13:8813-8826(1985).
RN   [9]
RP   SEQUENCE OF 3109-4563 FROM N.A.
RX   MEDLINE=85300528; PubMed=2994225;
RA   Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Urdea M.S.,
RA   Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M.,
RA   Priestley L.M., Robertson E., Rall L.B., Betsholtz C., Shows T.B.,
RA   Mahley R.W., Scott J.;
RT   "Human apolipoprotein B: structure of carboxyl-terminal domains, sites
RT   of gene expression, and chromosomal localization.";
RL   Science 230:37-43(1985).
RN   [10]
RP   SEQUENCE OF 1-291 FROM N.A.
RX   MEDLINE=86149325; PubMed=3513177;
RA   Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA   Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
RT   "Isolation of a cDNA clone encoding the amino-terminal region of human
RT   apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
```

```
RN   [11]
RP   SEQUENCE OF 1-1670 FROM N.A., AND VARIANT ILE-98.
RX   MEDLINE=86287319; PubMed=3461454;
RA   Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanaka M.,
RA   Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT   "Analysis of cDNA clones encoding the entire B-26 region of human
RT   apolipoprotein B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN   [12]
RP   PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX   MEDLINE=88018019; PubMed=3659919;
RA   Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA   Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
RA   Gotto A.M. Jr., Li W.-H., Chan L.;
RT   "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT   specific in-frame stop codon.";
RL   Science 238:363-366(1987).
RN   [13]
RP   DOMAINS.
RX   MEDLINE=87039351; PubMed=3773997;
RA   Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA   Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA   Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA   Levy-Wilson B., Scott J.;
RT   "Complete protein sequence and identification of structural domains of
RT   human apolipoprotein B.";
RL   Nature 323:734-738(1986).
RN   [14]
RP   DOMAINS.
RA   Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA   Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA   Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT   "Sequence, structure, receptor-binding domains and internal repeats of
RT   human apolipoprotein B-100.";
RL   Nature 323:738-742(1986).
RN   [15]
RP   CALCIUM-BINDING DATA.
RX   MEDLINE=86242245; PubMed=3087360;
RA   Dashti N., Lee D.M., Mok T.;
RT   "Apolipoprotein B is a calcium binding protein.";
RL   Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN   [16]
RP   PALMITOYLATION OF CYS-1112.
RX   MEDLINE=20143590; PubMed=10679026;
RA   Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT   "Palmitoylation of apolipoprotein B is required for proper
RT   intracellular sorting and transport of cholesteroyl esters and
RT   triglycerides.";
RL   Mol. Biol. Cell 11:721-734(2000).
RN   [17]
RP   VARIANT SER-4338.
RX   MEDLINE=91071750; PubMed=1979313;
RA   Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA   Cunny G., Cambien F., Roizes G.;
RT   "Detection by denaturing gradient gel electrophoresis of a new
RT   polymorphism in the apolipoprotein B gene.";
RL   Hum. Genet. 86:91-93(1990).
RN   [18]
RP   VARIANT FDB GLN-3527.
RX   MEDLINE=89098975; PubMed=2563166;
RA   Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA   McCarthy B.J.;
RT   "Association between a specific apolipoprotein B mutation and familial
RT   defective apolipoprotein B-100.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN   [19]
RP   VARIANT LEU-2739.
RX   MEDLINE=91016974; PubMed=2216805;
RA   Huang L.-S., Gavish D., Breslow J.L.;
RT   "Sequence polymorphism in the human apoB gene at position 8344.";
RL   Nucleic Acids Res. 18:5922-5922(1990).
RN   [20]
RP   VARIANT FDB CYS-3558.
RX   MEDLINE=95190020; PubMed=7883971;
RA   Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA   Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT   "Familial ligand-defective apolipoprotein B. Identification of a new
RT   mutation that decreases LDL receptor binding affinity.";
RL   J. Clin. Invest. 95:1225-1234(1995).
RN   [21]
RP   VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP   AND THR-4481.
RX   MEDLINE=97044521; PubMed=8889592;
RA   Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA   Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT   "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT   PCR-SSCP.";
RL   Hum. Mutat. 8:282-285(1996).
RN   [22]
RP   VARIANTS FDB GLN-3527 AND CYS-3558.
RX   MEDLINE=97403938; PubMed=9259199;
RA   Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA   Krempf M., Giraudet P., Junien C., Boileau C.;
RT   "Familial ligand-defective apolipoprotein B-100: simultaneous
RT   detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a French
RT   population.";
RL   Hum. Mutat. 10:160-163(1997).
RN   [23]
RP   VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP   AND ILE-3921.
RX   MEDLINE=98141125; PubMed=9490296;
RA   Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT   "Screening for mutations of the apolipoprotein B gene causing
RT   hypocholesterolemia.";
RL   Hum. Genet. 102:44-49(1998).
CC   -!- FUNCTION: Apolipoprotein B is a major protein constituent of
CC       chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo
CC       B-100 functions as a recognition signal for the cellular binding
CC       and internalization of LDL particles by the apoB/E receptor.
CC   -!- SUBCELLULAR LOCATION: Secreted.

  Query Match             84.0%;  Score 42;  DB 1;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 91;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 9
Q7Z600
ID   Q7Z600      PRELIMINARY;      PRT;  4563 AA.
AC   Q7Z600;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Apolipoprotein B (Including Ag(X) antigen).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA   Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY324608; AAP72970.1; -.
DR   GO; GO:0005319; F:lipid transporter activity; IEA.
DR   GO; GO:0006869; P:lipid transport; IEA.
DR   InterPro; IPR009454; DUF1081.
DR   InterPro; IPR001747; Lipid_transprt_N.
DR   Pfam; PF06448; DUF1081; 1.
DR   Pfam; PF01347; Vitellogenin_N; 1.
```

```
DR   SMART; SM00638; LPD_N; 1.
KW   Lipoprotein.
SQ   SEQUENCE   4563 AA;  515553 MW;  030B34167CEDC63C CRC64;

  Query Match             84.0%;  Score 42;  DB 2;  Length 4563;
  Best Local Similarity   90.0%;  Pred. No. 91;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              | ||||||||
Db       3385 TRLTRKRGLK 3394


RESULT 10
Q7YQQ0
ID   Q7YQQ0      PRELIMINARY;      PRT;   263 AA.
AC   Q7YQQ0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Procavia capensis (Cape hyrax) (Rock dassie).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Hyracoidea; Procaviidae; Procavia.
OX   NCBI_TaxID=9813;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548411; AAP97367.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     263    263
SQ   SEQUENCE   263 AA;  29532 MW;  536CF6149C1D062A CRC64;

  Query Match             80.0%;  Score 40;  DB 2;  Length 263;
  Best Local Similarity   80.0%;  Pred. No. 11;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              : ||||||||
Db        206 SSLTRKRGLK 215


RESULT 11
Q7M2U9
ID   Q7M2U9      PRELIMINARY;      PRT;   274 AA.
AC   Q7M2U9;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Apolipoprotein B-100 (Fragment).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90324804; PubMed=2373961;
RA   Law A., Scott J.;
RT   "A cross-species comparison of the apolipoprotein B domain that binds
RT   to the LDL receptor.";
RL   J. Lipid Res. 31:1109-1120(1990).
DR   PIR; A60950; A60950.
FT   NON_TER       1      1
FT   NON_TER     274    274
SQ   SEQUENCE   274 AA;  30505 MW;  CA1E1BE360AAB8F2 CRC64;

  Query Match             80.0%;  Score 40;  DB 2;  Length 274;
  Best Local Similarity   80.0%;  Pred. No. 12;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              : ||||||||
Db        221 SSLTRKRGLK 230


RESULT 12
Q7YQP9
ID   Q7YQP9      PRELIMINARY;      PRT;   304 AA.
AC   Q7YQP9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Echinops telfairi (Lesser hedgehog tenrec).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Insectivora; Tenrecidae; Tenrecinae; Echinops.
OX   NCBI_TaxID=9371;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548412; AAP97368.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     304    304
SQ   SEQUENCE   304 AA;  34264 MW;  468F4409260D6358 CRC64;

  Query Match             80.0%;  Score 40;  DB 2;  Length 304;
  Best Local Similarity   80.0%;  Pred. No. 13;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TDLTRKRGLK 10
              : ||||||||
Db        176 SSLTRKRGLK 185


RESULT 13
Q7YQN8
ID   Q7YQN8      PRELIMINARY;      PRT;   314 AA.
AC   Q7YQN8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Ictonyx striatus (striped polecat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
OC   Ictonyx.
OX   NCBI_TaxID=55050;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548425; AAP97381.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     314    314
SQ   SEQUENCE   314 AA;  34719 MW;  3E5B34E780F0039E CRC64;

  Query Match             80.0%;  Score 40;  DB 2;  Length 314;
  Best Local Similarity   80.0%;  Pred. No. 14;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 TDLTRKRGLK 10
              : ||||||||
Db        162 SSLTRKRGLK 171
```

RESULT 14
Q7YQP3
ID   Q7YQP3   PRELIMINARY;   PRT;   316 AA.
AC   Q7YQP3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Nandinia binotata (African palm civet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Viverridae; Nandiniinae;
OC   Nandinia.
OX   NCBI_TaxID=71115;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548420; AAP97376.1; -.
KW   Lipoprotein.
FT   NON_TER   1   1
FT   NON_TER   316   316
SQ   SEQUENCE   316 AA; 34540 MW; C04896B0E17562AE CRC64;

  Query Match   80.0%; Score 40; DB 2; Length 316;
  Best Local Similarity   80.0%; Pred. No. 14;
  Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

```
Qy          1 TDLTRKRGLK 10
              : ||||||||
Db        160 SSLTRKRGLK 169
```

RESULT 15
Q7YQN9
ID   Q7YQN9   PRELIMINARY;   PRT;   318 AA.
AC   Q7YQN9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Zalophus californianus (California sealion).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Pinnipedia; Otariidae; Zalophus.
OX   NCBI_TaxID=9704;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548424; AAP97380.1; -.
KW   Lipoprotein.
FT   NON_TER   1   1
FT   NON_TER   318   318
SQ   SEQUENCE   318 AA; 34888 MW; C04E7ECBA8E64C96 CRC64;

  Query Match   80.0%; Score 40; DB 2; Length 318;
  Best Local Similarity   80.0%; Pred. No. 14;
  Matches   8; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

```
Qy          1 TDLTRKRGLK 10
              : ||||||||
Db        162 SSLTRKRGLK 171
```

Search completed: January 13, 2005, 01:51:03
Job time : 79.0328 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:14:40 ; Search time 69.4918 Seconds
(without alignments)
46.460 Million cell updates/sec

Title: US-09-823-418-15
Perfect score: 44
Sequence: 1 TRLTRKGLK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	100.0	9	2	AAY30696	Aay30696 Apo-B100
2	41	93.2	9	2	AAY30694	Aay30694 Apo-B100
3	38	86.4	9	2	AAY30695	Aay30695 Apo-B100
4	34	77.3	601	6	ADA47994	Ada47994 Rice prot
5	34	77.3	798	6	ABU02213	Abu02213 S. pneumo
6	34	77.3	798	8	ADK46589	Adk46589 Streptoco
7	34	77.3	804	3	AAY55803	Aay55803 S. pneumo
8	33.5	76.1	10	2	AAY30690	Aay30690 Apo-B100
9	33.5	76.1	10	2	AAY30691	Aay30691 Apo-B100
10	33.5	76.1	11	2	AAW57205	Aaw57205 Apo B bin
11	33.5	76.1	13	2	AAW57207	Aaw57207 Apo B 100
12	33.5	76.1	15	2	AAW41261	Aaw41261 Apolipopr
13	33.5	76.1	15	2	AAW96892	Aaw96892 ApoB-100
14	33.5	76.1	20	6	ABJ37575	Abj37575 Heparin b
15	33.5	76.1	22	2	AAW57208	Aaw57208 Apo B 100
16	33.5	76.1	22	2	AAW57209	Aaw57209 Apo B 100
17	33.5	76.1	34	5	AAE14541	Aae14541 Human apo
18	33.5	76.1	36	2	AAW96876	Aaw96876 Nucleic a
19	33.5	76.1	37	2	AAW64587	Aaw64587 Human apo
20	33.5	76.1	51	2	AAW96845	Aaw96845 Nucleic a
21	33.5	76.1	343	4	ABB37687	Abb37687 Peptide #
22	33.5	76.1	343	4	ABG52504	Abg52504 Human liv
23	33.5	76.1	377	2	AAR72704	Aar72704 Human apo
24	33.5	76.1	377	2	AAR34031	Aar34031 Sequence
25	33.5	76.1	2463	8	ADJ57400	Adj57400 Human apo
26	33.5	76.1	3923	2	AAY31237	Aay31237 Human Apo
27	33.5	76.1	4536	2	AAW41262	Aaw41262 Apolipopr
28	33.5	76.1	4536	2	AAW96826	Aaw96826 Amino aci
29	33.5	76.1	4560	5	AAU98981	Aau98981 Human apo
30	33.5	76.1	4561	7	ADD48677	Add48677 Human Pro
31	33.5	76.1	4563	5	AAO15893	Aao15893 Human apo
32	33.5	76.1	4563	6	ABR40253	Abr40253 Human ali
33	33.5	76.1	4563	6	ABU79140	Abu79140 Apolipopr
34	33.5	76.1	4563	7	ADF43408	Adf43408 Apolipopr
35	33.5	76.1	4563	8	ADH18871	Adh18871 Human apo
36	33.5	76.1	4563	8	ADH18870	Adh18870 Human apo
37	33.5	76.1	4563	8	ADO33445	Ado33445 Human apo
38	33.5	76.1	4563	8	ADO33447	Ado33447 Human apo
39	33.5	76.1	4590	4	AAU33184	Aau33184 Novel hum
40	33	75.0	11	2	AAY30700	Aay30700 Apo-B100
41	33	75.0	11	2	AAY30698	Aay30698 Apo-B100
42	33	75.0	147	5	ABB49220	Abb49220 Listeria
43	33	75.0	150	7	ADC89104	Adc89104 Ribosomal
44	33	75.0	151	7	ADC88027	Adc88027 Ribosomal
45	33	75.0	154	7	ADM26985	Adm26985 Hyperther

ALIGNMENTS

RESULT 1
AAY30696
ID AAY30696 standard; peptide; 9 AA.
XX
AC AAY30696;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can

CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 44; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKGLK 9
|||||||||
Db 1 TRLTRKGLK 9

RESULT 2
AAY30694
ID AAY30694 standard; peptide; 9 AA.
XX
AC AAY30694;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 9 AA;

Query Match 93.2%; Score 41; DB 2; Length 9;
Best Local Similarity 88.9%; Pred. No. 1.7e+06;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKGLK 9
||||:|||
Db 1 TRLTRRGLK 9

RESULT 3
AAY30695
ID AAY30695 standard; peptide; 9 AA.
XX
AC AAY30695;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 9 AA;

Query Match 86.4%; Score 38; DB 2; Length 9;
Best Local Similarity 77.8%; Pred. No. 1.7e+06;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKGLK 9

```
                ||||::|||
Db          1 TRLTKRGLK 9


RESULT 4
ADA47994
ID   ADA47994 standard; protein; 601 AA.
XX
AC   ADA47994;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice protein conferring disease resistance in plants.
XX
KW   disease resistance; pathogen tolerance; plant pathogen; plant; rice.
XX
OS   Oryza sativa.
XX
PN   WO2003000906-A2.
XX
PD   03-JAN-2003.
XX
PF   21-JUN-2002; 2002WO-IB002453.
XX
PR   22-JUN-2001; 2001US-0300112P.
PR   26-SEP-2001; 2001US-0352277P.
PR   22-MAR-2002; 2002US-0366535P.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Glazebrook J,  Briggs S,  Cooper B,  Goff SA,  Moughamer T;
PI   Katagiri F,  Kreps J,  Provart N,  Ricke D,  Zhu T;
XX
DR   WPI; 2003-184052/18.
DR   N-PSDB; ADA47993.
XX
PT   New polynucleotide comprising a plant nucleotide sequence having an open
PT   reading frame that encodes a polypeptide associated with disease
PT   resistance, useful for conferring resistance or tolerance to a plant
PT   pathogen.
XX
PS   Claim 10; SEQ ID NO 64; 299pp; English.
XX
CC   The invention relates to a novel isolated polynucleotide comprising a
CC   plant nucleotide sequence having an open reading frame that encodes a
CC   polypeptide associated with disease resistance or its fragment having
CC   substantially the same activity as the full-length polypeptide. The
CC   polynucleotide of the invention is useful for conferring resistance or
CC   tolerance to a plant pathogen. The present sequence represents a protein
CC   conferring disease resistance used in the invention.
XX
SQ   Sequence 601 AA;

  Query Match             77.3%;  Score 34;  DB 6;  Length 601;
  Best Local Similarity   87.5%;  Pred. No. 2.2e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGL 8
              ||||| ||
Db        263 TRLTRSGL 270


RESULT 5
ABU02213
ID   ABU02213 standard; protein; 798 AA.
XX
AC   ABU02213;
XX
DT   23-OCT-2003  (revised)
DT   11-FEB-2003  (first entry)
XX
DE   S. pneumoniae type 4 strain protein from coding region #1791.
XX
KW   Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW   antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW   gene therapy; vaccine.
XX
OS   Streptococcus pneumoniae; type 4 strain.
XX
PN   WO200277021-A2.
XX
PD   03-OCT-2002.
XX
PF   27-MAR-2002; 2002WO-IB002163.
XX
PR   27-MAR-2001; 2001GB-00007658.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Masignani V,  Tettelin H,  Fraser C;
XX
DR   WPI; 2003-040579/03.
DR   N-PSDB; ABX07503.
XX
PT   New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT   useful as medicaments for treating or preventing a disease or infection
PT   due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT   ear infection.
XX
PS   Claim 1; SEQ ID NO 3582; 56pp; English.
XX
CC   The invention relates to a protein comprising or having at least 50%
CC   identity to any of the 2469 amino acid sequences, identified in the
CC   specification (available on a computer readable format), or its fragment,
CC   expressed from 2469 of 2489 identified DNA coding regions from the
CC   Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC   ABS56454. Also included are an antibody which binds one of the proteins,
CC   treating a patient by administering the protein, DNA or antibody (in a
CC   composition), a kit comprising first and second primers, which are the
CC   nucleic acid cited above or fragments between nucleotides 8-100 of a
CC   sequence not defined in the specification, for amplifying a target
CC   sequence contained within a Streptococcus nucleic acid sequence, where
CC   the first primer is substantially complementary to the target sequence
CC   and the second primer is substantially complementary to the complement of
CC   the target sequence, and where the parts of the primers having
CC   substantial complementarity define the termini of the target sequence to
CC   be amplified, assay comprising contacting a test compound with the
CC   protein, and determining whether the test compound binds to the protein
CC   and a Streptococcus pneumoniae bacterium, where one or more genes
CC   encoding the proteins has been rendered inactive. The proteins, nucleic
CC   acid molecules, antibody and compositions are useful as medicaments for
CC   treating or preventing a disease or infection due to streptococcus
CC   bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC   media or ear infection. They are also useful in developing vaccines,
CC   diagnostics and antibiotics. The methods are useful for identifying
CC   immunodominant proteins. The present sequence is one of the 2469 proteins
CC   expressed by the identified coding regions from the genomic sequence.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 798 AA;

  Query Match             77.3%;  Score 34;  DB 6;  Length 798;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKGL 8
              |||||||
Db        156 RLTRKGL 162


RESULT 6
ADK46589
```

ID ADK46589 standard; protein; 798 AA.
XX
AC ADK46589;
XX
DT 20-MAY-2004 (first entry)
XX
DE Streptococcus pneumoniae protein, Seq ID No 3104.
XX
KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
PN US6699703-B1.
XX
PD 02-MAR-2004.
XX
PF 26-MAY-2000; 2000US-00583110.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
DR WPI; 2004-212399/20.
DR N-PSDB; ADK43928.
XX
PT New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
PS Disclosure; SEQ ID NO 3104; 301pp; English.
XX
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 798 AA;

Query Match 77.3%; Score 34; DB 8; Length 798;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RLTRKGL 8
|||||||
Db 156 RLTRKGL 162

RESULT 7
AAY55803
ID AAY55803 standard; protein; 804 AA.
XX
AC AAY55803;
XX
DT 28-FEB-2000 (first entry)
XX
DE S. pneumoniae priA polypeptide.
XX
KW priA polypeptide; microbial disease; vaccine; microbial infection;
KW Streptococcus pneumoniae; antibacterial.
XX
OS Streptococcus pneumoniae.
XX
PN WO9961453-A2.
XX
PD 02-DEC-1999.
XX
PF 22-APR-1999; 99WO-US008771.
XX
PR 27-APR-1998; 98US-00067091.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Mcdevitt D, Shilling L, Warren RL, St John A;
XX
DR WPI; 2000-062670/05.
DR N-PSDB; AAZ39569, AAZ39572.
XX
PT New isolated priA polypeptides, useful for screening antibacterial
PT compounds.
XX
PS Claim 1; Page 4-5; 68pp; English.
XX
CC This represents a S. pneumoniae priA polypeptide. The priA polypeptides
CC and polynucleotides are useful for the treatment of microbial diseases
CC (especially in the form of vaccines) and the methods are useful for
CC identifying agonists and antagonists. The polypeptides are also useful
CC for relating to diagnostic assays for detecting diseases associated with
CC microbial infections (especially infections by Streptococcus pneumoniae)
CC and conditions associated with such infections and assays for detecting
CC priA expression or activity. The polypeptides are useful in the discovery
CC and development of antibacterial compounds. The encoded protein upon
CC expression can be used as a target for screening of antibacterial drugs
XX
SQ Sequence 804 AA;

Query Match 77.3%; Score 34; DB 3; Length 804;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RLTRKGL 8
|||||||
Db 162 RLTRKGL 168

RESULT 8
AAY30690
ID AAY30690 standard; peptide; 10 AA.
XX
AC AAY30690;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing

PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 76.1%; Score 33.5; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRK-GLK 9
|||||| |||
Db 1 TRLTRKEGLK 10

RESULT 9
AAY30691
ID AAY30691 standard; peptide; 10 AA.
XX
AC AAY30691;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 76.1%; Score 33.5; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRK-GLK 9
|||||| |||
Db 1 TRLTRKDGLK 10

RESULT 10
AAW57205
ID AAW57205 standard; peptide; 11 AA.
XX
AC AAW57205;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B binding site peptide 2.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 12; Page 52; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B binding site
CC peptide which can be used as a component of a non-naturally occurring,
CC receptor-competent low density lipoprotein (LDL) particle of the present
CC invention. The LDL particle comprises at least 1 peptide component that
CC has at least 1 binding site for an apo B protein receptor and at least 1
CC lipophilic substituent. Also described in the invention are peptides
CC containing an apo B binding sequence with at least 70% identity with
CC sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their dimers. Non-
CC naturally occurring, receptor-competent LDL particles are useful as: (i)
CC drug-targeting vectors for delivering anticancer drugs to cancer cells

```
CC   that express an apo B protein receptor, and (ii) additives for cell
CC   culture media especially as growth supplements. Non-naturally occurring,
CC   receptor-competent LDL particles do not require the complete apo B
CC   sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 11 AA;

  Query Match             76.1%;  Score 33.5;  DB 2;  Length 11;
  Best Local Similarity   90.0%;  Pred. No. 4.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRK-GLK 9
              |||||| |||
Db          2 TRLTRKRGLK 11


RESULT 11
AAW57207
ID   AAW57207 standard; peptide; 13 AA.
XX
AC   AAW57207;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide B.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 13 AA;

  Query Match             76.1%;  Score 33.5;  DB 2;  Length 13;
  Best Local Similarity   90.0%;  Pred. No. 5;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRK-GLK 9
              |||||| |||
Db          3 TRLTRKRGLK 12


RESULT 12
AAW41261
ID   AAW41261 standard; peptide; 15 AA.
XX
AC   AAW41261;
XX
DT   19-MAY-1998  (first entry)
XX
DE   Apolipoprotein B-100 fragment.
XX
KW   Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW   thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
KW   angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW   prothrombinase complex.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9743311-A1.
XX
PD   20-NOV-1997.
XX
PF   09-MAY-1997;   97WO-GB001255.
XX
PR   09-MAY-1996;   96GB-00009702.
XX
PA   (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI   Bruckdorfer KR,  Ettelaie C;
XX
DR   WPI; 1998-008798/01.
XX
PT   Peptide fragments of apo:lipo:protein B-100 with anticoagulant activity -
PT   used for treating or preventing coagulation, inhibiting angiogenesis,
PT   cell differentiation and apoptosis.
XX
PS   Disclosure; Page 22; 60pp; English.
XX
CC   This sequence is an example of the peptide of the invention. It has the
CC   formula (I), or their variants with one or more internal deletions,
CC   insertions or substitutions, while retaining anti-coagulant properties of
CC   apolipoprotein B-100 (apoB-100). Z1-KAQ-X1-KKNKHRHS-X2-T-Z2 (I) X1 = S or
CC   Y; X2 = T or I; Z1 = the N terminus of the peptide, or 1-47 amino acids
CC   (aa); Z2 = the C terminus of the peptide, a terminal amide group or 1-77
CC   aa. Compositions containing the peptide are used for simultaneous,
CC   separate or sequential treatment of cancer, particularly to prevent
CC   metastatic spread. They are also used to inhibit thromboplastin-mediated
CC   processes, specifically to prevent or reduce blood coagulation (e.g.
CC   during or after surgery or in cases of heart attack, stroke etc.) and to
CC   inhibit angiogenesis, cellular differentiation or apoptosis. KRAD-14,
CC   which is active as such or as part of a 98-aa peptide, inhibits
CC   activation of the prothrombinase complex; and prevents activation of
CC   factor VII on the surface of thromboplastin and of platelets by thrombin.
CC   It binds to the residues 58-66 of thromboplastin. Since (I) are much
CC   smaller than apoB-100, they act more quickly
XX
SQ   Sequence 15 AA;

  Query Match             76.1%;  Score 33.5;  DB 2;  Length 15;
  Best Local Similarity   90.0%;  Pred. No. 5.8;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;
```

```
Qy          1 TRLTRK-GLK 9
              |||||| |||
Db          1 TRLTRKRGLK 10


RESULT 13
AAW96892
ID   AAW96892 standard; peptide; 15 AA.
XX
AC   AAW96892;
XX
DT   22-APR-1999  (first entry)
XX
DE   ApoB-100 nuclear localisation signal sequence, residues 3353-3367.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO9856938-A1.
XX
PD   17-DEC-1998.
XX
PF   10-JUN-1998;   98WO-US011927.
XX
PR   13-JUN-1997;   97US-00874807.
PR   14-MAY-1998;   98US-00079030.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Guevara JG,  Hoogeveen RC,  Moore JP;
XX
DR   WPI; 1999-070331/06.
XX
PT   Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT   used for delivering nucleic acid to cells for gene therapy and antisense
PT   treatment.
XX
PS   Claim 19; Fig 13D; 293pp; English.
XX
CC   AAW96878-97 represent nuclear localisation signal sequence derived from
CC   human apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein
CC   component of very-low density lipoproteins (VLDL), intermediate density
CC   lipoprotein (IDL), low density lipoproteins ((LDL) and lipoprotein a. The
CC   present sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 15 AA;

  Query Match             76.1%;  Score 33.5;  DB 2;  Length 15;
  Best Local Similarity   90.0%;  Pred. No. 5.8;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRK-GLK 9
              |||||| |||
Db          6 TRLTRKRGLK 15


RESULT 14
ABJ37575
ID   ABJ37575 standard; peptide; 20 AA.
XX
AC   ABJ37575;
XX
DT   10-MAY-2003  (first entry)
XX
DE   Heparin binding peptide sequence #28.
XX
KW   Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW   cardiovascular; circulatory; ligand; sulphated; sulphonated; tumour;
KW   rheumatoid arthritis; hypoxia; diabetic retinopathy; heparin.
XX
OS   Unidentified.
XX
PN   WO2003007689-A2.
XX
PD   30-JAN-2003.
XX
PF   22-JUL-2002; 2002WO-US023419.
XX
PR   20-JUL-2001; 2001US-0306726P.
XX
PA   (ETHZ-) ETH ZUERICH.
PA   (UYZU-) UNIV ZURICH.
XX
PI   Hubbell JA,  Schoenmakers R,  Maynard HD;
XX
DR   WPI; 2003-300420/29.
XX
PT   Use of a ligand comprising of at least one sulfated or sulfonated amino
PT   acid for the treatment of e.g. tumors, rheumatoid arthritis, diabetic
PT   retinopathy and hypoxia.
XX
PS   Disclosure; Fig 2; 79pp; English.
XX
CC   The invention relates to a novel ligand for binding a target biomolecule,
CC   which comprises a peptide having at least one sulphated or sulphonated
CC   amino acid and at least one amino acid chosen from neutral and positively
CC   charged amino acids. The novel ligands can be used for the treatment of
CC   e.g. tumours, rheumatoid arthritis, diabetic retinopathy and hypoxia.
CC   This sequence represents a heparin binding peptide relating to the
CC   invention
XX
SQ   Sequence 20 AA;

  Query Match             76.1%;  Score 33.5;  DB 6;  Length 20;
  Best Local Similarity   90.0%;  Pred. No. 7.8;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRK-GLK 9
              |||||| |||
Db          7 TRLTRKRGLK 16


RESULT 15
AAW57208
ID   AAW57208 standard; peptide; 22 AA.
XX
AC   AAW57208;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide C.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
FT   Modified-site   22
FT                   /note= "attached to cholesterol"
XX
PN   WO9813385-A2.
```

XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 76.1%; Score 33.5; DB 2; Length 22;
Best Local Similarity 90.0%; Pred. No. 8.7;
Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy          1 TRLTRK-GLK 9
              |||||| |||
Db          7 TRLTRKRGLK 16
```

Search completed: January 13, 2005, 01:43:03
Job time : 70.6585 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:15:45 ; Search time 70.2295 Seconds
(without alignments)
73.735 Million cell updates/sec

Title: US-09-823-418-15
Perfect score: 44
Sequence: 1 TRLTRKGLK 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	84.1	423	2	Q7WSQ9	Q7wsq9 arthrobacte
2	36	81.8	871	2	O64892	O64892 ananas como
3	35	79.5	453	2	Q6BJK6	Q6bjk6 debaryomyce
4	34	77.3	138	1	RBFA_CHRVO	Q7ny12 chromobacte
5	34	77.3	172	2	Q9XBV5	Q9xbv5 mycobacteri
6	34	77.3	376	2	Q72E67	Q72e67 desulfovibr
7	34	77.3	376	2	AAS95192	Aas95192 desulfovi
8	34	77.3	451	1	SYS_PYRAE	Q8ztp4 pyrobaculum
9	34	77.3	457	2	Q9FKT8	Q9fkt8 arabidopsis
10	34	77.3	482	2	Q6Z5C5	Q6z5c5 oryza sativ
11	34	77.3	482	2	BAD10124	Bad10124 oryza sat
12	34	77.3	493	2	Q8TJF9	Q8tjf9 methanosarc
13	34	77.3	798	2	Q8DNR6	Q8dnr6 streptococc
14	34	77.3	798	2	Q97PA5	Q97pa5 streptococc
15	34	77.3	1230	2	Q7G6A5	Q7g6a5 oryza sativ
16	34	77.3	1230	2	Q8RU68	Q8ru68 oryza sativ
17	33.5	76.1	414	2	Q7YQR5	Q7yqr5 aotus vocif
18	33.5	76.1	596	2	Q28473	Q28473 macaca fasc
19	33.5	76.1	3262	2	Q13788	Q13788 homo sapien
20	33.5	76.1	4563	1	APB_HUMAN	P04114 homo sapien
21	33.5	76.1	4563	2	Q7Z600	Q7z600 homo sapien
22	33	75.0	102	2	Q8R2E8	Q8r2e8 rattus norv
23	33	75.0	137	2	Q7MI23	Q7mi23 vibrio vuln
24	33	75.0	137	2	Q8DBU7	Q8dbu7 vibrio vuln
25	33	75.0	147	1	NDK_LISMO	Q8y5x4 listeria mo
26	33	75.0	147	2	Q71Y86	Q71y86 listeria mo
27	33	75.0	147	2	AAT04728	Aat04728 listeria
28	33	75.0	150	1	RS13_BRUPA	P62299 brugia paha
29	33	75.0	150	1	RS13_WUCBA	P62300 wuchereria
30	33	75.0	151	1	NDK_ARCFU	O29491 archaeoglob
31	33	75.0	151	2	Q17274	Q17274 brugia paha
32	33	75.0	154	1	NDK_METKA	Q8tv10 methanopyru
33	33	75.0	157	2	Q9L533	Q91533 vibrio chol
34	33	75.0	174	2	Q9KKZ5	Q9kkz5 vibrio chol
35	33	75.0	306	2	Q87WZ9	Q87wz9 pseudomonas
36	33	75.0	335	2	Q8PNK9	Q8pnk9 xanthomonas
37	33	75.0	470	1	ROCC_BACSU	P39636 bacillus su
38	33	75.0	503	2	Q7P2I3	Q7p2i3 fusobacteri
39	33	75.0	506	2	Q8REI1	Q8rei1 fusobacteri
40	33	75.0	592	2	Q6PAP4	Q6pap4 mus musculu
41	33	75.0	592	2	AAH57340	Aah57340 mus muscu
42	33	75.0	592	2	AAH60177	Aah60177 mus muscu
43	33	75.0	716	2	Q80YS3	Q80ys3 mus musculu
44	33	75.0	761	1	PHT1_MOUSE	Q9qz09 mus musculu
45	33	75.0	762	1	PHT1_HUMAN	Q9ums5 homo sapien

ALIGNMENTS

RESULT 1
Q7WSQ9
ID Q7WSQ9 PRELIMINARY; PRT; 423 AA.
AC Q7WSQ9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative transporter protein.
OS Arthrobacter ilicis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=43665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rue61a;
RX MEDLINE=22753791; PubMed=12730200;
RA Parschat K., Hauer B., Kappl R., Kraft R., Huettermann J., Fetzner S.;
RT "Gene Cluster of Arthrobacter ilicis R.61a Involved in the Degradation
RT of Quinaldine to Anthranilate. Characterization and Functional
RT Expression of the Quinaldine 4-oxidase qoxLMS Genes.";
RL J. Biol. Chem. 278:27483-27494(2003).
DR EMBL; AJ537472; CAD61041.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR PROSITE; PS50850; MFS; 1.
SQ SEQUENCE 423 AA; 43696 MW; BB11CBADA85DF241 CRC64;

Query Match 84.1%; Score 37; DB 2; Length 423;
Best Local Similarity 77.8%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TRLTRKGLK 9
              ||||::|||
Db        207 TRLTKQGLK 215
```

RESULT 2
O64892
ID O64892 PRELIMINARY; PRT; 871 AA.
AC O64892;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Ananas comosus (Pineapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Bromeliaceae;
OC Ananas.
OX NCBI_TaxID=4615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98418625; PubMed=9747853;

```
RA   Thomson K.G., Thomas J.E., Dietzgen R.G.;
RT   "Retrotransposon-like sequences integrated into the genome of
RT   pineapple, Ananas comosus.";
RL   Plant Mol. Biol. 38:461-465(1998).
DR   EMBL; Y12432; CAA73042.1; -.
DR   PIR; T07863; T07863.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00078; RVT; 1.
KW   Polyprotein; RNA-directed DNA polymerase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     871    871
SQ   SEQUENCE   871 AA;  100048 MW;  EDFFD016E08952FC CRC64;

  Query Match             81.8%;  Score 36;  DB 2;  Length 871;
  Best Local Similarity   77.8%;  Pred. No. 72;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGLK 9
              |||| ||:|
Db        272 TRLTHKGVK 280


RESULT 3
Q6BJK6
ID   Q6BJK6      PRELIMINARY;      PRT;   453 AA.
AC   Q6BJK6;
DT   01-OCT-2004 (TrEMBLrel. 28, Created)
DT   01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similar to CA1599|IPF11452 Candida albicans IPF11452 of unknown
DE   function.
GN   ORFNames=DEHA0G02068g;
OS   Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX   NCBI_TaxID=4959;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RG   GENOLEVURES;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382139; CAG90062.1; -.
SQ   SEQUENCE   453 AA;  50757 MW;  DD4AF51CF5D956F5 CRC64;

  Query Match             79.5%;  Score 35;  DB 2;  Length 453;
  Best Local Similarity   77.8%;  Pred. No. 58;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGLK 9
              :|||||| |
Db        368 SRLTRKGTK 376


RESULT 4
RBFA_CHRVO
ID   RBFA_CHRVO     STANDARD;      PRT;   138 AA.
AC   Q7NY12;
DT   29-MAR-2004 (Rel. 43, Created)
DT   29-MAR-2004 (Rel. 43, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Ribosome-binding factor A.
GN   Name=rbfA; OrderedLocusNames=CV1463;
OS   Chromobacterium violaceum.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Chromobacterium.
OX   NCBI_TaxID=536;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12472 / DSM 30191;
RX   MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA   Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA   Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA   Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
RA   Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA   Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
RA   Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,
RA   Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA   Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
RA   Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA   Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA   Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
RA   Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA   Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA   Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA   Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
RA   di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA   Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA   Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA   Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA   Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA   Santos E.B.P., Santos F.R., Schneider M.P.C., Seuanez H.N.,
RA   Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA   Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA   Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA   Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT   "The complete genome sequence of Chromobacterium violaceum reveals
RT   remarkable and exploitable bacterial adaptability.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
CC   -!- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC       with 30S subunits that are part of 70S ribosomes or polysomes).
CC       Essential for efficient processing of 16S rRNA. May interact with
CC       the 5'terminal helix region of 16S rRNA (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the rbfA family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016915; AAQ59138.2; -.
DR   HAMAP; MF_00003; -; 1.
DR   InterPro; IPR000238; Rib_bind_factA.
DR   Pfam; PF02033; RBFA; 1.
DR   ProDom; PD007327; Rib_bind_factA; 1.
```

```
DR   TIGRFAMs; TIGR00082; rbfA; 1.
DR   PROSITE; PS01319; RBFA; 1.
KW   Complete proteome; rRNA processing.
SQ   SEQUENCE   138 AA;  15387 MW;  E84750D86390272C CRC64;

  Query Match             77.3%;  Score 34;  DB 1;  Length 138;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LTRKGLK 9
              |||||||
Db         25 LTRKGLK 31


RESULT 5
Q9XBV5
ID   Q9XBV5      PRELIMINARY;      PRT;   172 AA.
AC   Q9XBV5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mycobacterium smegmatis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1772;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=mc2 155;
RA   Bardou F., Martinez R., Puech V., Bleiber G., Prod'hom G., Daffe M.,
RA   Telenti A.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155062; AAD42390.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   172 AA;  18622 MW;  C4DFA92D0182E682 CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 172;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LTRKGLK 9
              |||||||
Db        132 LTRKGLK 138


RESULT 6
Q72E67
ID   Q72E67      PRELIMINARY;      PRT;   376 AA.
AC   Q72E67;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Amino acid ABC transporter, periplasmic-binding protein.
GN   OrderedLocusNames=DVU0712;
OS   Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS   8303).
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC   Desulfovibrionaceae; Desulfovibrio.
OX   NCBI_TaxID=882;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=15077118; DOI=10.1038/nbt959;
RA   Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA   Kolonay J.F., Eisen J.A., Ward N.L., Methe B.A., Brinkac L.M.,
RA   Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA   Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA   Peterson J.D., Davidsen T.M., Zafar N., Zhou L., Radune D.,
RA   Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA   Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT   "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT   Desulfovibrio vulgaris Hildenborough.";
RL   Nat. Biotechnol. 22:554-559(2004).
DR   EMBL; AE017311; AAS95192.1; -.
DR   TIGR; DVU0712; -.
DR   InterPro; IPR001828; ANF_receptor.
DR   InterPro; IPR000709; Leu_Ile_Val_bind.
DR   Pfam; PF01094; ANF_receptor; 1.
DR   PRINTS; PR00337; LEUILEVALBP.
KW   Complete proteome.
SQ   SEQUENCE   376 AA;  40203 MW;  14F0AFE97CC2B79B CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 376;
  Best Local Similarity   87.5%;  Pred. No. 80;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKGLK 9
              ||| ||||
Db        130 RLTEKGLK 137


RESULT 7
AAS95192
ID   AAS95192    PRELIMINARY;      PRT;   376 AA.
AC   AAS95192;
DT   27-APR-2004 (TrEMBLrel. 27, Created)
DT   27-APR-2004 (TrEMBLrel. 27, Last sequence update)
DT   11-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE   Amino acid ABC transporter, periplasmic-binding protein.
GN   DVU0712.
OS   Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS   8303).
OC   Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC   Desulfovibrionaceae; Desulfovibrio.
OX   NCBI_TaxID=882;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=15077118;
RA   Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA   Kolonay J.F., Eisen J.A., Ward N., Methe B.A., Brinkac L.M.,
RA   Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA   Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA   Peterson J.D., Davidsen T.M., Zafar N., Zhou L., Radune D.,
RA   Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Utterback T.R.,
RA   Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT   "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT   Desulfovibrio vulgaris Hildenborough.";
RL   Nat. Biotechnol. 22:554-559(2004).
DR   EMBL; AE017311; AAS95192.1; -.
DR   TIGR; DVU0712; -.
SQ   SEQUENCE   376 AA;  40203 MW;  14F0AFE97CC2B79B CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 376;
  Best Local Similarity   87.5%;  Pred. No. 80;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKGLK 9
              ||| ||||
Db        130 RLTEKGLK 137


RESULT 8
SYS_PYRAE
ID   SYS_PYRAE      STANDARD;      PRT;   451 AA.
AC   Q8ZTP4;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Seryl-tRNA synthetase (EC 6.1.1.11) (Serine--tRNA ligase) (SerRS).
GN   Name=serS; OrderedLocusNames=PAE3158;
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
```

```
RX   PubMed=11792869; DOI=10.1073/pnas.241636498;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + L-serine + tRNA(Ser) = AMP + diphosphate
CC       + L-seryl-tRNA(Ser).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE009914; AAL64715.1; -.
DR   HSSP; P34945; 1SER.
DR   HAMAP; MF_00176; -; 1.
DR   InterPro; IPR002314; tRNA-synt_2b.
DR   InterPro; IPR002317; tRNA-synt_ser.
DR   InterPro; IPR010978; tRNA_binding_arm.
DR   InterPro; IPR006195; tRNA_ligase_II.
DR   Pfam; PF02403; Seryl_tRNA_N; 1.
DR   Pfam; PF00587; tRNA-synt_2b; 1.
DR   PRINTS; PR00981; TRNASYNTHSER.
DR   TIGRFAMs; TIGR00414; serS; 1.
DR   PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
KW   Aminoacyl-tRNA synthetase; ATP-binding; Complete proteome; Ligase;
KW   Protein biosynthesis.
SQ   SEQUENCE   451 AA;  52030 MW;  985B4C826E75D505 CRC64;

  Query Match             77.3%;  Score 34;  DB 1;  Length 451;
  Best Local Similarity   75.0%;  Pred. No. 98;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKGLK 9
              |:||||:|
Db        383 RVTRKGMK 390


RESULT 9
Q9FKT8
ID   Q9FKT8      PRELIMINARY;      PRT;   457 AA.
AC   Q9FKT8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Similarity to cytochrome oxidase assembly factor (Hypothetical protein
DE   At5g56090/MDA7_15) (Hypothetical protein At5g56090).
GN   Name=At5g56090/MDA7_15; Synonyms=At5g56090;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98344145; PubMed=9679202;
RA   Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT   features of the regions of 1,381,565 bp covered by twenty one
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:131-145(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA   Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA   Hayashizaki Y., Shinozaki K.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA   Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
RA   Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA   Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA   Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011476; BAB09291.1; -.
DR   EMBL; AK117496; BAC42159.1; -.
DR   EMBL; BT004976; AAO50509.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006461; P:protein complex assembly; IEA.
DR   InterPro; IPR003780; COX15_CtaA.
DR   Pfam; PF02628; COX15-CtaA; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   457 AA;  50127 MW;  6E7A0DE6457E2D1E CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 457;
  Best Local Similarity   87.5%;  Pred. No. 99;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGL 8
              ||||| ||
Db        120 TRLTRSGL 127


RESULT 10
Q6Z5C5
ID   Q6Z5C5      PRELIMINARY;      PRT;   482 AA.
AC   Q6Z5C5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Putative cytochrome c oxidase subunit 15(COX15) homolog isoform
DE   1.
GN   Name=B1142B04.27;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sasaki T., Matsumoto T., Katayose Y.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005148; BAD10124.1; -.
DR   InterPro; IPR003780; COX15_CtaA.
DR   Pfam; PF02628; COX15-CtaA; 1.
SQ   SEQUENCE   482 AA;  51446 MW;  DC46727EC38CEAD5 CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 482;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGL 8
              ||||| ||
Db        147 TRLTRSGL 154


RESULT 11
BAD10124
ID   BAD10124    PRELIMINARY;      PRT;   482 AA.
AC   BAD10124;
DT   02-MAR-2004 (TrEMBLrel. 27, Created)
DT   02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   Putative cytochrome c oxidase subunit 15(COX15) homolog isoform
DE   1.
GN   B1142B04.27.
```

```
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza; Oryza sativa.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Katayose Y.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
RT   clone:B1142B04.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005148; BAD10124.1; -.
SQ   SEQUENCE   482 AA;  51446 MW;  DC46727EC38CEAD5 CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 482;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGL 8
              ||||| ||
Db        147 TRLTRSGL 154


RESULT 12
Q8TJF9
ID   Q8TJF9      PRELIMINARY;      PRT;   493 AA.
AC   Q8TJF9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Terminase.
GN   OrderedLocusNames=MA3826;
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE011094; AAM07177.1; -.
DR   InterPro; IPR006517; DUF_phage_C.
DR   TIGRFAMs; TIGR01630; psiM2_ORF9; 1.
KW   Complete proteome.
SQ   SEQUENCE   493 AA;  56892 MW;  0170B5011675FD0C CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 493;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LTRKGLK 9
              |||||||
Db        473 LTRKGLK 479


RESULT 13
Q8DNR6
ID   Q8DNR6      PRELIMINARY;      PRT;   798 AA.
AC   Q8DNR6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Primosomal replication factor Y.
GN   Name=priA; OrderedLocusNames=spr1581;
OS   Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=171101;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21429245; PubMed=11544234;
RA   Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burgett S.,
RA   DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA   Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA   LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA   McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA   Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA   Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA   Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA   Glass J.I.;
RT   "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL   J. Bacteriol. 183:5709-5717(2001).
DR   EMBL; AE008525; AAL00384.1; -.
DR   PIR; C98069; C98069.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0015668; F:type III site-specific deoxyribonuclease ac. . .; IEA.
DR   GO; GO:0009307; P:DNA restriction; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR005259; PriA.
DR   InterPro; IPR006935; ResIII.
DR   Pfam; PF00271; Helicase_C; 1.
DR   Pfam; PF04851; ResIII; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   TIGRFAMs; TIGR00595; priA; 1.
KW   ATP-binding; Complete proteome; Helicase; Hydrolase.
SQ   SEQUENCE   798 AA;  90036 MW;  DD0C4EEA5B269962 CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 798;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKGL 8
              |||||||
Db        156 RLTRKGL 162


RESULT 14
Q97PA5
ID   Q97PA5      PRELIMINARY;      PRT;   798 AA.
AC   Q97PA5;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Primosomal protein N'.
GN   OrderedLocusNames=SP1736;
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC BAA-334 / TIGR4;
RX   MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA   Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA   Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA   Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
```

```
RA   Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA   Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA   McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA   Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA   Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT   "Complete genome sequence of a virulent isolate of Streptococcus
RT   pneumoniae.";
RL   Science 293:498-506(2001).
DR   EMBL; AE007466; AAK75812.1; -.
DR   PIR; C95202; C95202.
DR   TIGR; SP1736; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR005259; PriA.
DR   Pfam; PF00271; Helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   TIGRFAMs; TIGR00595; priA; 1.
KW   ATP-binding; Complete proteome; Helicase; Hydrolase.
SQ   SEQUENCE   798 AA;  90028 MW;  FED84D4A8BB9B198 CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 798;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKGL 8
              |||||||
Db        156 RLTRKGL 162


RESULT 15
Q7G6A5
ID   Q7G6A5      PRELIMINARY;      PRT;  1230 AA.
AC   Q7G6A5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Putative 22 kDa kafirin cluster; Ty3-Gypsy type.
GN   ORFNames=OSJNAb0075K12.33;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
RA   Rambo T., Saski C., Henry D., Oates R., Simmons J.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC122148; AAM48279.1; -.
DR   EMBL; AE017082; AAP53268.1; -.
DR   InterPro; IPR009007; Pept_Aspartic.
DR   InterPro; IPR001969; Pept_Asp_AS.
DR   InterPro; IPR005162; Retrotrans_gag.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF03732; Retrotrans_gag; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00078; RVT; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   RNA-directed DNA polymerase; Transferase.
SQ   SEQUENCE   1230 AA;  141401 MW;  28B6664530A45846 CRC64;

  Query Match             77.3%;  Score 34;  DB 2;  Length 1230;
  Best Local Similarity   66.7%;  Pred. No. 3e+02;
  Matches    6;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TRLTRKGLK 9
              ||||:||::
Db        706 TRLTQKGVR 714


Search completed: January 13, 2005, 01:51:05
Job time : 72.2295 secs
```

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:14:40 ; Search time 84.9344 Seconds
(without alignments)
46.460 Million cell updates/sec

Title: US-09-823-418-17
Perfect score: 54
Sequence: 1 TRLTRKERGLK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	100.0	11	2	AAY30698	Aay30698 Apo-B100
2	51	94.4	11	2	AAY30700	Aay30700 Apo-B100
3	46	85.2	11	2	AAY30697	Aay30697 Apo-B100
4	44	81.5	11	2	AAY30699	Aay30699 Apo-B100
5	38.5	71.3	10	2	AAY30690	Aay30690 Apo-B100
6	38.5	71.3	10	2	AAY30682	Aay30682 Apo-B100
7	38.5	71.3	11	2	AAW57205	Aaw57205 Apo B bin
8	38.5	71.3	13	2	AAW57207	Aaw57207 Apo B 100
9	38.5	71.3	15	2	AAW41261	Aaw41261 Apolipopr
10	38.5	71.3	15	2	AAW96892	Aaw96892 ApoB-100
11	38.5	71.3	20	6	ABJ37575	Abj37575 Heparin b
12	38.5	71.3	22	2	AAW57208	Aaw57208 Apo B 100
13	38.5	71.3	22	2	AAW57209	Aaw57209 Apo B 100
14	38.5	71.3	34	5	AAE14541	Aae14541 Human apo
15	38.5	71.3	36	2	AAW96876	Aaw96876 Nucleic a
16	38.5	71.3	37	2	AAW64587	Aaw64587 Human apo
17	38.5	71.3	51	2	AAW96845	Aaw96845 Nucleic a
18	38.5	71.3	343	4	ABB37687	Abb37687 Peptide #
19	38.5	71.3	343	4	ABG52504	Abg52504 Human liv
20	38.5	71.3	377	2	AAR72704	Aar72704 Human apo
21	38.5	71.3	377	2	AAR34031	Aar34031 Sequence
22	38.5	71.3	2463	8	ADJ57400	Adj57400 Human apo
23	38.5	71.3	3923	2	AAY31237	Aay31237 Human Apo
24	38.5	71.3	4536	2	AAW41262	Aaw41262 Apolipopr
25	38.5	71.3	4536	2	AAW96826	Aaw96826 Amino aci
26	38.5	71.3	4560	5	AAU98981	Aau98981 Human apo
27	38.5	71.3	4561	7	ADD48677	Add48677 Human Pro
28	38.5	71.3	4563	5	AAO15893	Aao15893 Human apo
29	38.5	71.3	4563	6	ABR40253	Abr40253 Human ali
30	38.5	71.3	4563	6	ABU79140	Abu79140 Apolipopr
31	38.5	71.3	4563	7	ADF43408	Adf43408 Apolipopr
32	38.5	71.3	4563	8	ADH18871	Adh18871 Human apo
33	38.5	71.3	4563	8	ADH18870	Adh18870 Human apo
34	38.5	71.3	4563	8	ADO33445	Ado33445 Human apo
35	38.5	71.3	4563	8	ADO33447	Ado33447 Human apo
36	38.5	71.3	4590	4	AAU33184	Aau33184 Novel hum
37	38	70.4	97	4	AAU21930	Aau21930 Human car
38	38	70.4	97	7	ADE45898	Ade45898 Human car
39	38	70.4	270	4	ABG14088	Abg14088 Novel hum
40	38	70.4	1023	4	ABG06061	Abg06061 Novel hum
41	37	68.5	488	5	ABB93444	Abb93444 Herbicida
42	36	66.7	1095	5	ABP52170	Abp52170 Mouse pot
43	36	66.7	1095	7	ADD37509	Add37509 Mouse pho
44	36	66.7	1095	8	ADI27988	Adi27988 Murine pr
45	35.5	65.7	10	2	AAY30683	Aay30683 Apo-B100

ALIGNMENTS

RESULT 1
AAY30698
ID AAY30698 standard; peptide; 11 AA.
XX
AC AAY30698;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can

CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 54; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKERGLK 11
||||||||||||
Db 1 TRLTRKERGLK 11

RESULT 2
AAY30700
ID AAY30700 standard; peptide; 11 AA.
XX
AC AAY30700;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 58; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 11 AA;

Query Match 94.4%; Score 51; DB 2; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.016;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKERGLK 11
||||||:||||
Db 1 TRLTRKDRGLK 11

RESULT 3
AAY30697
ID AAY30697 standard; peptide; 11 AA.
XX
AC AAY30697;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 2; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.13;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKERGLK 11

```
              |||||::||||
Db          1 TRLTREKRGLK 11


RESULT 4
AAY30699
ID   AAY30699 standard; peptide; 11 AA.
XX
AC   AAY30699;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW   Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW   low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946598-A1.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US004805.
XX
PR   10-MAR-1998;   98US-0077618P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Innerarity TL,  Boren JOS;
XX
DR   WPI; 1999-551509/46.
XX
PT   Identifying compounds which affect binding of low density lipoprotein
PT   with proteoglycan, used for, e.g. obtaining compounds for reducing
PT   atherosclerosis.
XX
PS   Claim 17; Page 58; 70pp; English.
XX
CC   AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC   receptor mutations. They were created to identify compounds which
CC   modulate atherosclerosis. The peptides are derived from amino acids 3358
CC   to 3367 of apoB100. The method comprises detecting compounds which affect
CC   low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC   can be used for identifying compounds which disrupt LDL-PG binding
CC   without inhibiting LDL receptor binding. Such compounds can be used to
CC   reduce or prevent the formation of atherosclerotic lesions and prevent
CC   atherosclerosis. The transgenic non-human animals and mammals which
CC   express human apo-B100 can be used as an in vivo model system for the
CC   study of atherosclerosis, and in vivo assay methods for identifying
CC   compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 11 AA;

  Query Match             81.5%;  Score 44;  DB 2;  Length 11;
  Best Local Similarity   81.8%;  Pred. No. 0.3;
  Matches    9;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKERGLK 11
              ||||| :||||
Db          1 TRLTRDKRGLK 11


RESULT 5
AAY30690
ID   AAY30690 standard; peptide; 10 AA.
XX
AC   AAY30690;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW   Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW   low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946598-A1.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US004805.
XX
PR   10-MAR-1998;   98US-0077618P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Innerarity TL,  Boren JOS;
XX
DR   WPI; 1999-551509/46.
XX
PT   Identifying compounds which affect binding of low density lipoprotein
PT   with proteoglycan, used for, e.g. obtaining compounds for reducing
PT   atherosclerosis.
XX
PS   Claim 17; Page 57; 70pp; English.
XX
CC   AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC   receptor mutations. They were created to identify compounds which
CC   modulate atherosclerosis. The peptides are derived from amino acids 3358
CC   to 3367 of apoB100. The method comprises detecting compounds which affect
CC   low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC   can be used for identifying compounds which disrupt LDL-PG binding
CC   without inhibiting LDL receptor binding. Such compounds can be used to
CC   reduce or prevent the formation of atherosclerotic lesions and prevent
CC   atherosclerosis. The transgenic non-human animals and mammals which
CC   express human apo-B100 can be used as an in vivo model system for the
CC   study of atherosclerosis, and in vivo assay methods for identifying
CC   compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 10 AA;

  Query Match             71.3%;  Score 38.5;  DB 2;  Length 10;
  Best Local Similarity   90.9%;  Pred. No. 2.8;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              ||||||| |||
Db          1 TRLTRKE-GLK 10


RESULT 6
AAY30682
ID   AAY30682 standard; peptide; 10 AA.
XX
AC   AAY30682;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
```

KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 10 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 10;
Best Local Similarity 90.9%; Pred. No. 2.8;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
|||||| |||||
Db 1 TRLTR-ERGLK 10

RESULT 7
AAW57205
ID AAW57205 standard; peptide; 11 AA.
XX
AC AAW57205;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B binding site peptide 2.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 12; Page 52; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B binding site
CC peptide which can be used as a component of a non-naturally occurring,
CC receptor-competent low density lipoprotein (LDL) particle of the present
CC invention. The LDL particle comprises at least 1 peptide component that
CC has at least 1 binding site for an apo B protein receptor and at least 1
CC lipophilic substituent. Also described in the invention are peptides
CC containing an apo B binding sequence with at least 70% identity with
CC sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their dimers. Non-
CC naturally occurring, receptor-competent LDL particles are useful as: (i)
CC drug-targeting vectors for delivering anticancer drugs to cancer cells
CC that express an apo B protein receptor, and (ii) additives for cell
CC culture media especially as growth supplements. Non-naturally occurring,
CC receptor-competent LDL particles do not require the complete apo B
CC sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 11 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 11;
Best Local Similarity 90.9%; Pred. No. 3.1;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
|||||| ||||
Db 2 TRLTRK-RGLK 11

RESULT 8
AAW57207
ID AAW57207 standard; peptide; 13 AA.
XX
AC AAW57207;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide B.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.

XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 13 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 13;
Best Local Similarity 90.9%; Pred. No. 3.6;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
|||||| ||||
Db 3 TRLTRK-RGLK 12

RESULT 9
AAW41261
ID AAW41261 standard; peptide; 15 AA.
XX
AC AAW41261;
XX
DT 19-MAY-1998 (first entry)
XX
DE Apolipoprotein B-100 fragment.
XX
KW Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
KW angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW prothrombinase complex.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9743311-A1.
XX
PD 20-NOV-1997.
XX
PF 09-MAY-1997; 97WO-GB001255.
XX
PR 09-MAY-1996; 96GB-00009702.
XX
PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI Bruckdorfer KR, Ettelaie C;
XX
DR WPI; 1998-008798/01.
XX
PT Peptide fragments of apo:lipo:protein B-100 with anticoagulant activity -
PT used for treating or preventing coagulation, inhibiting angiogenesis,
PT cell differentiation and apoptosis.
XX
PS Disclosure; Page 22; 60pp; English.
XX
CC This sequence is an example of the peptide of the invention. It has the
CC formula (I), or their variants with one or more internal deletions,
CC insertions or substitutions, while retaining anti-coagulant properties of
CC apolipoprotein B-100 (apoB-100). Z1-KAQ-X1-KKNKHRHS-X2-T-Z2 (I) X1 = S or
CC Y; X2 = T or I; Z1 = the N terminus of the peptide, or 1-47 amino acids
CC (aa); Z2 = the C terminus of the peptide, a terminal amide group or 1-77
CC aa. Compositions containing the peptide are used for simultaneous,
CC separate or sequential treatment of cancer, particularly to prevent
CC metastatic spread. They are also used to inhibit thromboplastin-mediated
CC processes, specifically to prevent or reduce blood coagulation (e.g.
CC during or after surgery or in cases of heart attack, stroke etc.) and to
CC inhibit angiogenesis, cellular differentiation or apoptosis. KRAD-14,
CC which is active as such or as part of a 98-aa peptide, inhibits
CC activation of the prothrombinase complex; and prevents activation of
CC factor VII on the surface of thromboplastin and of platelets by thrombin.
CC It binds to the residues 58-66 of thromboplastin. Since (I) are much
CC smaller than apoB-100, they act more quickly
XX
SQ Sequence 15 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 15;
Best Local Similarity 90.9%; Pred. No. 4.2;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
|||||| ||||
Db 1 TRLTRK-RGLK 10

RESULT 10
AAW96892
ID AAW96892 standard; peptide; 15 AA.
XX
AC AAW96892;
XX
DT 22-APR-1999 (first entry)
XX
DE ApoB-100 nuclear localisation signal sequence, residues 3353-3367.
XX
KW Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Claim 19; Fig 13D; 293pp; English.
XX
CC AAW96878-97 represent nuclear localisation signal sequence derived from

CC human apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein
CC component of very-low density lipoproteins (VLDL), intermediate density
CC lipoprotein (IDL), low density lipoproteins ((LDL) and lipoprotein a. The
CC present sequence can be used in the composition of the invention. The
CC specification describes a composition that comprises LDL and
CC apolipoproteins for the binding and in vivo transport of nucleic acids.
CC The composition is used to deliver nucleic acids to eukaryotic cells, in
CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC molecule (or ribozyme). Specifically they are used for gene therapy of
CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC fibrosis and arteriosclerosis
XX
SQ Sequence 15 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 15;
Best Local Similarity 90.9%; Pred. No. 4.2;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
||||||| ||||
Db 6 TRLTRK-RGLK 15

RESULT 11
ABJ37575
ID ABJ37575 standard; peptide; 20 AA.
XX
AC ABJ37575;
XX
DT 10-MAY-2003 (first entry)
XX
DE Heparin binding peptide sequence #28.
XX
KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW cardiovascular; circulatory; ligand; sulphated; sulphonated; tumour;
KW rheumatoid arthritis; hypoxia; diabetic retinopathy; heparin.
XX
OS Unidentified.
XX
PN WO2003007689-A2.
XX
PD 30-JAN-2003.
XX
PF 22-JUL-2002; 2002WO-US023419.
XX
PR 20-JUL-2001; 2001US-0306726P.
XX
PA (ETHZ-) ETH ZUERICH.
PA (UYZU-) UNIV ZURICH.
XX
PI Hubbell JA, Schoenmakers R, Maynard HD;
XX
DR WPI; 2003-300420/29.
XX
PT Use of a ligand comprising of at least one sulfated or sulfonated amino
PT acid for the treatment of e.g. tumors, rheumatoid arthritis, diabetic
PT retinopathy and hypoxia.
XX
PS Disclosure; Fig 2; 79pp; English.
XX
CC The invention relates to a novel ligand for binding a target biomolecule,
CC which comprises a peptide having at least one sulphated or sulphonated
CC amino acid and at least one amino acid chosen from neutral and positively
CC charged amino acids. The novel ligands can be used for the treatment of
CC e.g. tumours, rheumatoid arthritis, diabetic retinopathy and hypoxia.
CC This sequence represents a heparin binding peptide relating to the
CC invention
XX
SQ Sequence 20 AA;

Query Match 71.3%; Score 38.5; DB 6; Length 20;
Best Local Similarity 90.9%; Pred. No. 5.5;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
||||||| ||||
Db 7 TRLTRK-RGLK 16

RESULT 12
AAW57208
ID AAW57208 standard; peptide; 22 AA.
XX
AC AAW57208;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide C.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
FT Modified-site 22
FT /note= "attached to cholesterol"
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 22 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 22;
Best Local Similarity 90.9%; Pred. No. 6.1;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
||||||| ||||
Db 7 TRLTRK-RGLK 16

```
RESULT 13
AAW57209
ID   AAW57209 standard; peptide; 22 AA.
XX
AC   AAW57209;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide D.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 22 AA;

  Query Match             71.3%;  Score 38.5;  DB 2;  Length 22;
  Best Local Similarity   90.9%;  Pred. No. 6.1;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db          7 TRLTRK-RGLK 16


RESULT 14
AAE14541
ID   AAE14541 standard; peptide; 34 AA.
XX
AC   AAE14541;
XX
DT   17-MAY-2002  (first entry)
XX
DE   Human apoB-100 derived peptide p62.
XX
KW   Oxidised low density lipoprotein; oxLDL; immunoassay; periaortitis;
KW   cardiovascular disease; coronary heart disease; pre-eclampsia;
KW   non-insulin-dependent diabetes; endothelial dysfunction; human; apoB-100;
KW   peptide p62.
XX
OS   Homo sapiens.
XX
PN   WO200206314-A2.
XX
PD   24-JAN-2002.
XX
PF   18-JUL-2001; 2001WO-GB003212.
XX
PR   18-JUL-2000; 2000GB-00017641.
XX
PA   (ARKT-) ARK THERAPEUTICS LTD.
XX
PI   Narvanen O,  Yla-Herttuala S;
XX
DR   WPI; 2002-179777/23.
XX
PT   New peptide useful in enzyme immunoassays for detecting oxidized low
PT   density lipoprotein which is a marker of coronary heart disease and other
PT   cardiovascular diseases, has affinity for oxidized low density
PT   lipoprotein.
XX
PS   Claim 6; Page 5; 21pp; English.
XX
CC   The invention relates to peptides having affinity for oxidised low
CC   density lipoprotein (oxLDL), in cyclised or multimeric form. The peptide
CC   is useful in an immunoassay to determine the presence, and optionally,
CC   the amount of antibodies in a sample, having affinity for oxLDL.
CC   Preferably immobilised peptide is useful for measuring the amount of
CC   autoantibodies for oxLDL in a sample, especially a serum or plasma sample
CC   from a patient for evaluating the risk of coronary heart diseases, other
CC   cardiovascular diseases, and several other disorders such as
CC   periaortitis, pre-eclampsia, non-insulin-dependent diabetes and
CC   endothelial dysfunction. The peptide of the invention is stable, can be
CC   synthesised easily without the need to isolate proteins from a patient's
CC   blood, and has a long half-life. The present sequence is human apoB-100
CC   derived peptide p62 used in the invention
XX
SQ   Sequence 34 AA;

  Query Match             71.3%;  Score 38.5;  DB 5;  Length 34;
  Best Local Similarity   90.9%;  Pred. No. 9.4;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db         25 TRLTRK-RGLK 34


RESULT 15
AAW96876
ID   AAW96876 standard; peptide; 36 AA.
XX
AC   AAW96876;
XX
DT   22-APR-1999  (first entry)
XX
DE   Nucleic acid binding domain from apoB-100, residues 3348-3390.
XX
KW   Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW   apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW   nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW   non-small cell lung carcinoma; diabetes; arteriosclerosis.
```

XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense
PT treatment.
XX
PS Claim 16; Fig 12C; 293pp; English.
XX
CC AAW96827-77 represent nucleic acid binding domains derived from human
CC apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein component
CC of very-low density lipoproteins (VLDL), intermediate density lipoprotein
CC (IDL), low density lipoproteins ((LDL) and lipoprotein a. The present
CC sequence can be used in the composition of the invention. The
CC specification describes a composition that comprises LDL and
CC apolipoproteins for the binding and in vivo transport of nucleic acids.
CC The composition is used to deliver nucleic acids to eukaryotic cells, in
CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC molecule (or ribozyme). Specifically they are used for gene therapy of
CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC fibrosis and arteriosclerosis
XX
SQ Sequence 36 AA;

Query Match 71.3%; Score 38.5; DB 2; Length 36;
Best Local Similarity 90.9%; Pred. No. 9.9;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db         11 TRLTRK-RGLK 20
```

Search completed: January 13, 2005, 01:43:04
Job time : 86.1011 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:30:05 ; Search time 15.8689 Seconds
(without alignments)
66.696 Million cell updates/sec

Title: US-09-823-418-17
Perfect score: 54
Sequence: 1 TRLTRKERGLK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	77.8	2073	1	BWASBE	bimE protein - Eme
2	38.5	71.3	596	2	S32802	apolipoprotein B -
3	38.5	71.3	4563	1	LPHUB	apolipoprotein B-1
4	37	68.5	488	2	T49903	glucosyltransferas
5	36	66.7	248	2	H84008	hypothetical prote
6	36	66.7	250	2	D98082	hypothetical prote
7	36	66.7	264	2	AI2112	hypothetical prote
8	36	66.7	405	2	C72305	transposase, IS605
9	35	64.8	100	2	A36950	urease (EC 3.5.1.5
10	35	64.8	346	2	S49963	hypothetical prote
11	35	64.8	807	2	AC1031	hypothetical prote
12	34.5	63.9	269	2	C60950	apolipoprotein B-1
13	34.5	63.9	779	2	JH0102	apolipoprotein B -
14	34	63.0	221	2	B64188	arginine transport
15	34	63.0	250	2	A69843	hypothetical prote
16	34	63.0	608	2	A46312	gag polyprotein -
17	33	61.1	191	2	T25791	hypothetical prote
18	33	61.1	246	2	T38787	hypothetical prote
19	33	61.1	284	2	G82319	DnaJ-related prote
20	33	61.1	305	1	NKVLHH	core antigen - her
21	33	61.1	332	2	T13447	hypothetical prote
22	33	61.1	541	2	AF0547	propionate catabol
23	33	61.1	605	2	T04197	hypothetical prote
24	33	61.1	643	2	AF2471	penicillin-binding
25	33	61.1	710	2	T46589	ropy-2 protein [im
26	32.5	60.2	275	2	E60950	apolipoprotein B-1
27	32	59.3	100	2	S47102	urease (EC 3.5.1.5
28	32	59.3	132	2	G75409	hypothetical prote
29	32	59.3	150	2	H72645	hypothetical prote
30	32	59.3	202	2	G84502	hypothetical prote
31	32	59.3	220	2	A75287	response regulator
32	32	59.3	250	2	F95218	hypothetical prote
33	32	59.3	266	2	AB1827	hypothetical prote
34	32	59.3	301	2	AF2223	heterodisulfide re
35	32	59.3	345	2	T48758	hypothetical prote
36	32	59.3	398	2	S40752	hypothetical prote
37	32	59.3	429	2	E83723	hypothetical prote
38	32	59.3	441	2	JQ2191	nucleocapsid prote
39	32	59.3	458	2	D70410	cytosolic axial fi
40	32	59.3	627	2	T00484	hypothetical prote
41	32	59.3	637	2	JH0611	glutamate-cysteine
42	32	59.3	637	2	A35015	glutamate-cysteine
43	32	59.3	771	2	T29177	hypothetical prote
44	32	59.3	798	2	AI2053	competence protein
45	32	59.3	844	2	S05988	translation elonga

ALIGNMENTS

RESULT 1
BWASBE
bimE protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C;Accession: A37879
R;Engle, D.B.; Osmani, S.A.; Osmani, A.H.; Rosborough, S.; Xiang, X.; Morris, N.R.
J. Biol. Chem. 265, 16132-16137, 1990
A;Title: A negative regulator of mitosis in Aspergillus is a putative membrane-spannin
A;Reference number: A37879; MUID:90375468; PMID:1697851
A;Accession: A37879
A;Molecule type: mRNA
A;Residues: 1-2073 <ENG>
A;Cross-references: UNIPROT:P24686; GB:M59705; GB:J05607; NID:g168026; PIDN:AAA51478.1,
A;Note: in addition to three predicted transmembrane domains, there are several potent
asein kinase, and one sequence that resembles a nuclear localization signal
C;Comment: This protein is part of a regulatory pathway that includes the nimA protein
ter mitosis and prevent them from leaving mitosis.
C;Genetics:
A;Gene: bimE
C;Superfamily: bimE protein
C;Keywords: cell cycle control; mitosis; transmembrane protein
F;1623-1643/Domain: transmembrane #status predicted <TM1>
F;1685-1703/Domain: transmembrane #status predicted <TM2>
F;1746-1764/Domain: transmembrane #status predicted <TM3>

Query Match 77.8%; Score 42; DB 1; Length 2073;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TRLTRKERGL 10
              |||||::|||
Db        832 TRLTRRKRGL 841
```

RESULT 2
S32802
apolipoprotein B - crab-eating macaque (fragment)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S32802
R;Pape, M.E.; Castle, C.K.; Murray, R.W.; Funk, G.M.; Hunt, C.E.; Marotti, K.R.; Melch
Biochim. Biophys. Acta 1086, 326-334, 1991
A;Title: Apo B metabolism in the cynomolgus monkey: evidence for post-transcriptional
A;Reference number: S32802; MUID:92075708; PMID:1742325
A;Accession: S32802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-596 <PAP>
A;Cross-references: UNIPROT:Q28473; EMBL:X15737; NID:g38047; PIDN:CAA33755.1; PID:g930
C;Superfamily: apolipoprotein B

```
  Query Match             71.3%;  Score 38.5;  DB 2;  Length 596;
  Best Local Similarity   90.9%;  Pred. No. 21;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db        226 TRLTRK-RGLK 235
```

RESULT 3
LPHUB
apolipoprotein B-100 precursor - human
N;Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
C;Species: Homo sapiens (man)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: A27850; A25679; A25263; A25267; A25266; A24320; A24684; A23817; A25774; A26
4452; I61909; I59510; I39474; I39469; I84624; I37179; PS0058
R;Ludwig, E.H.; Blackhart, B.D.; Pierotti, V.R.; Caiati, L.; Fortier, C.; Knott, T.; Scc
DNA 6, 363-372, 1987
A;Title: DNA sequence of the human apolipoprotein B gene.
A;Reference number: A27850; MUID:88003974; PMID:3652907
A;Accession: A27850
A;Molecule type: DNA
A;Residues: 1-617,'A',619-1929,'F',1931-3318,'D',3320-3426,'T',3428-3431,'Q',3433-3731,'
A;Cross-references: UNIPROT:P04114; UNIPROT:P78482; UNIPROT:P78479; UNIPROT:Q9UMN0; UNIP
R;Cladaras, C.; Hadzopoulou-Cladaras, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A;Title: The complete sequence and structural analysis of human apolipoprotein B-100: re
A;Reference number: A91058; MUID:87161758; PMID:3030729
A;Accession: A25679
A;Molecule type: mRNA
A;Residues: 1-11,15-2539,'S',2541-3823,'R',3825-4563 <CLA>
A;Note: 1109-Asp was also found
R;Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McCa
Nucleic Acids Res. 14, 7501-7503, 1986
A;Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A;Reference number: A93639; MUID:87016385; PMID:3763409
A;Accession: A25263
A;Molecule type: mRNA
A;Residues: 1-272,'N',274-617,'A',619-1217,'E',1219-2091,'V',2093-2364,'T',2366-2679,'Q'
A;Cross-references: GB:X04506; NID:g34330; PIDN:CAA28191.1; PID:g34331
R;Law, S.W.; Grant, S.M.; Higuchi, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer Jr
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A;Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A;Reference number: A94134; MUID:87041416; PMID:3464946
A;Accession: A25267
A;Molecule type: mRNA
A;Residues: 1-617,'A',619-703,'P',705-792,'R',794-1270,'S',1272-1866,'G',1868-2036,'N',2
4189-4220,'M',4222-4563 <LAW>
A;Note: the codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4188, and
R;Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Gotto Jr., A.M
J. Biol. Chem. 261, 12918-12921, 1986
A;Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A;Reference number: A92556; MUID:87008488; PMID:3759943
A;Accession: A25266
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-328,'V',330-644,'I',646-918,'P',920-3318,'D',3320-3426,'T',3428-
9-4132,'G',4134-4180,'E',4182-4563 <CHE>
A;Cross-references: GB:J02610; NID:g178803; PIDN:AAA35549.1; PID:g178804
A;Note: a total of 2366 residues were confirmed by direct sequencing of tryptic peptides
R;Protter, A.A.; Hardman, D.A.; Sato, K.Y.; Schilling, J.W.; Yamanaka, M.; Hort, Y.J.; H
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A;Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein
A;Reference number: A24320; MUID:86287319; PMID:3461454
A;Accession: A24320
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-617,'A',619-941,'YYIWSLPPKP',951-1138,'PTGRLPNCFSNGLICYSLWLHSFQE
A;Cross-references: GB:M14081; NID:g178795; PIDN:AAA51752.1; PID:g553189
R;Law, S.W.; Lackner, K.J.; Hospattankar, A.V.; Anchors, J.M.; Sakaguchi, A.Y.; Naylor,
Proc. Natl. Acad. Sci. U.S.A. 82, 8340-8344, 1985
A;Title: Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of
A;Reference number: A24684; MUID:86094221; PMID:3001697
A;Accession: A24684
A;Molecule type: mRNA
A;Residues: 485-617,'A',619-1044 <LA2>
A;Cross-references: GB:M12480; NID:g178791; PIDN:AAA51751.1; PID:g178792
R;Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; K
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A;Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipop
A;Reference number: A94088; MUID:86149325; PMID:3513177
A;Accession: A23817
A;Molecule type: mRNA
A;Residues: 1-291 <PRO>
A;Cross-references: GB:M12681; NID:g178797; PIDN:AAA51753.1; PID:g178798
R;Deeb, S.S.; Motulsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A;Title: A partial cDNA clone for human apolipoprotein B.
A;Reference number: A25774; MUID:85270450; PMID:3860836
A;Accession: A25774
A;Molecule type: mRNA
A;Residues: 709-791,'SSSWKAASHGCPHSAGD',810-906 <DEE>
A;Cross-references: GB:K03175; NID:g178821; PIDN:AAA51759.1; PID:g178822
R;Carlsson, P.; Darnfors, C.; Olofsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A;Title: Analysis of the human apolipoprotein B gene; complete structure of the B-74 re
A;Reference number: A91565; MUID:87191999; PMID:2883086
A;Accession: A26533
A;Molecule type: mRNA
A;Residues: 1282-2721,2742-3290,'L',3292-3336,'N',3338-3948,'F',3950-3963,'Y',3965-4180
A;Cross-references: GB:M15421; NID:g178817; PIDN:AAA51758.1; PID:g178818
R;Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Sato, K.Y.; Lau, K.; Yaman
Biochemistry 26, 5478-5486, 1987
A;Title: Structural comparison of human apolipoproteins B-48 and B-100.
A;Reference number: A29671; MUID:88050832; PMID:3676265
A;Accession: A29671
A;Molecule type: mRNA
A;Residues: 1671-2323,'PYW',2327-2352,'H',2354-2398 <HAR>
A;Cross-references: GB:M17367; NID:g178731; PIDN:AAA51741.1; PID:g178732
R;Shoulders, C.C.; Myant, N.B.; Sidoli, A.; Rodriguez, J.C.; Cortese, C.; Baralle, F.E.
Atherosclerosis 58, 277-289, 1985
A;Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for more than o
A;Reference number: A90084; MUID:86130855; PMID:3841481
A;Accession: A29287
A;Molecule type: mRNA
A;Residues: 3846-4298 <SHO>
R;Pfitzner, R.; Wagener, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986
A;Title: Isolation, expression and characterization of a human apolipoprotein B 100-spe
A;Reference number: A25572; MUID:87076044; PMID:3024665
A;Accession: A25572
A;Molecule type: mRNA
A;Residues: 4219-4337,'S',4339-4563 <PFI>
A;Cross-references: GB:M36676
R;Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow, J.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269, 1985
A;Reference number: A24738; MUID:86042646; PMID:2932736
A;Accession: A24738
A;Molecule type: mRNA
A;Residues: 'N',3729-3731,'I',3733-3875,'A',3877-3948,'F',3950-3963,'Y',3965-3982,'S',3
A;Cross-references: GB:M12413; NID:g178735; PIDN:AAA51742.1; PID:g178736
R;Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.; Silberman, S.R.; Ca
Science 238, 363-366, 1987
A;Title: Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific i
A;Reference number: A40133; MUID:88018019; PMID:3659919
A;Accession: B40133
A;Molecule type: mRNA
A;Residues: 2165-2179 <CH1>
A;Cross-references: GB:M18036; NID:g178799; PIDN:AAA51754.1; PID:g178800
A;Note: this mRNA includes the stop codon of the organ-specific mRNA for apo48
A;Accession: A40133
A;Molecule type: protein
A;Residues: 51-75;101-110;129-139;158-174;197-207;276-287;298-304;306-314;526-532;538-5
36;1486-1498;1537-1556;1563-1572;1601-1610;1647-1661;1697-1724;1770-1781;1859-1897;1968
A;Note: these fragments were derived from apo48
R;Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987
```
```

A;Title: Carboxyl terminal analysis of human B-48 protein confirms the novel mechanism p
A;Reference number: A28002; MUID:88106542; PMID:3426612
A;Accession: A28002
A;Molecule type: mRNA
A;Residues: 2129-2179,2181-2235 <HA2>
A;Cross-references: GB:M18471
A;Experimental source: intestine
A;Note: this mRNA from intestine includes a stop codon created by RNA editing in place c
R;Mehrabian, M.; Schumaker, V.N.; Fareed, G.C.; West, R.; Johnson, D.F.; Kirchgessner, T
Nucleic Acids Res. 13, 6937-6953, 1985
A;Title: Human apolipoprotein B: identification of cDNA clones and characterization of m
A;Reference number: A24269; MUID:86041888; PMID:3903660
A;Accession: A24269
A;Molecule type: mRNA
A;Residues: 3056-3159 <MEH>
A;Cross-references: GB:X03045; NID:g28783; PIDN:CAA26850.1; PID:g929609
R;Hospattankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer Jr., H.B.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A;Title: Identification of a novel in-frame translational stop codon in human intestine
A;Reference number: A29659; MUID:88049670; PMID:2445342
A;Accession: A29659
A;Molecule type: mRNA
A;Residues: 2169-2179 <HOS>
A;Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A;Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human intest
ch encodes the 250K apoB-48, CAA encoding 2180-Gln is substituted by the stop codon TAA,
R;Yang, C.; Kim, T.W.; Weng, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
A;Title: Isolation and characterization of sulfhydryl and disulfide peptides of human ap
A;Reference number: A35783; MUID:90319144; PMID:2115173
A;Contents: disulfide bonds
A;Accession: A35783
A;Molecule type: protein
A;Residues: 28-41;76-97,'I',99-100;175-193;206-215;239-249;259-266;357-399;455-490;512-5

A;Note: cysteines at positions 1112, 1422, 1505, 1662, 3761, 3917, and 4217 have free su
R;LeBoeuf, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
FEBS Lett. 170, 105-108, 1984
A;Title: Human apolipoprotein B: partial amino acid sequence.
A;Reference number: A22006; MUID:84208786; PMID:6373369
A;Accession: A22006
A;Molecule type: protein
A;Residues: 873-892,'K',894-896 <LE1>
A;Accession: B22006
A;Molecule type: protein
A;Residues: 3113,'L',3115-3130,'R',3132-3133,'P',3135-3136,'R' <LE2>
R;Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Caiati, L.; Onasch, M.A.; Wallis, S.C.;
J. Biol. Chem. 261, 15364-15367, 1986
A;Title: Structure of the human apolipoprotein B gene.
A;Reference number: A92564; MUID:87057153; PMID:2946672
A;Contents: annotation; gene structure
R;Wagener, R.; Pfitzner, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
A;Title: Studies on the organization of the human apolipoprotein B 100 gene.
A;Reference number: A90715; MUID:87271140; PMID:2886136
A;Contents: annotation; gene structure
R;Weisgraber, K.H.; Rall Jr., S.C.
J. Biol. Chem. 262, 11097-11103, 1987
A;Title: Human apolipoprotein B-100 heparin-binding sites.
A;Reference number: A92605; MUID:87280197; PMID:3301850
A;Contents: annotation; heparin binding and disulfide bond
R;Dashti, N.; Lee, D.M.; Mok, T.
Biochem. Biophys. Res. Commun. 137, 493-499, 1986
A;Title: Apolipoprotein B is a calcium binding protein.
A;Reference number: A90125; MUID:86242245; PMID:3087360
A;Contents: annotation; calcium binding
R;Carlsson, P.; Olofsson, S.O.; Bondjers, G.; Darnfors, C.; Wiklund, O.; Bjursell, G.
Nucleic Acids Res. 13, 8813-8826, 1985
A;Title: Molecular cloning of human apolipoprotein B cDNA.
A;Reference number: I37178; MUID:86093680; PMID:3841204
A;Accession: I37180

Query Match 71.3%; Score 38.5; DB 1; Length 4563;
Best Local Similarity 90.9%; Pred. No. 1.3e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKERGLK 11
||||||.||||
Db 3385 TRLTRK-RGLK 3394

RESULT 4
T49903
glucosyltransferase-like protein - Arabidopsis thaliana
N;Alternate names: protein T24H18.60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49903
R;Bevan, M.; Robben, J.; Grymonprez, B.; Volckaert, G.; Bancroft, I.; Mewes, H.W.; Rudd
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25024
A;Accession: T49903
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-488 <BEV>
A;Cross-references: UNIPROT:Q9LXV0; EMBL:AL353013; GSPDB:GN00063; ATSP:T24H18.60
A;Experimental source: cultivar Columbia; BAC clone T24H18
C;Genetics:
A;Gene: ATSP:T24H18.60
A;Map position: 5
C;Superfamily: flavonol O3-glucosyltransferase

Query Match 68.5%; Score 37; DB 2; Length 488;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RLTRKERGL 10
|:|| ||||
Db 344 RITRSERGL 352

RESULT 5
H84008
hypothetical protein BH2872 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: H84008
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H84008
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: UNIPROT:Q9K8Y0; GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH2872
C;Superfamily: Bacillus subtilis hypothetical protein yjbA

Query Match 66.7%; Score 36; DB 2; Length 248;
Best Local Similarity 88.9%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LTRKERGLK 11
||||||.||
Db 155 LTRKERQLK 163

RESULT 6
D98082
hypothetical protein fecE [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: D98082

```
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: D98082
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <KUR>
A;Cross-references: UNIPROT:Q8DNJ3; GB:AE007317; PIDN:AAL00489.1; PID:g15459361; GSPDB:G
C;Genetics:
A;Gene: fecE
C;Superfamily: inner membrane protein malK; ATP-binding cassette homology

  Query Match             66.7%;  Score 36;  DB 2;  Length 250;
  Best Local Similarity   60.0%;  Pred. No. 28;
  Matches    6;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TRLTRKERGL 10
              :|||:|:||:
Db         47 SRLTKKDRGV 56


RESULT 7
AI2112
hypothetical protein all2456 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AI2112
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2112
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-264 <KUR>
A;Cross-references: UNIPROT:Q8YU98; GB:BA000019; PIDN:BAB74155.1; PID:g17131548; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2456
C;Superfamily: conserved hypothetical protein HI0188

  Query Match             66.7%;  Score 36;  DB 2;  Length 264;
  Best Local Similarity   87.5%;  Pred. No. 30;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 LTRKERGL 10
              |||:||||
Db        124 LTRRERGL 131


RESULT 8
C72305
transposase, IS605-TnpB family - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: C72305
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: C72305
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-405 <ARN>
A;Cross-references: UNIPROT:Q9X0D4; GB:AE001764; GB:AE000512; NID:g4981561; PIDN:AAD3612
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1044
C;Superfamily: hypothetical protein b1432

  Query Match             66.7%;  Score 36;  DB 2;  Length 405;
  Best Local Similarity   70.0%;  Pred. No. 44;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGLK 11
              ||:||:|| |
Db        227 RLSRKQRGSK 236


RESULT 9
A36950
urease (EC 3.5.1.5) 11k chain - Bacillus sp. (strain TB-90)
N;Alternate names: ureA protein
C;Species: Bacillus sp.
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 09-Jul-2004
C;Accession: A36950
R;Maeda, M.; Hidaka, M.; Nakamura, A.; Masaki, H.; Uozumi, T.
J. Bacteriol. 176, 432-442, 1994
A;Title: Cloning, sequencing, and expression of thermophilic Bacillus sp. strain TB-90
A;Reference number: A36950; MUID:94117379; PMID:8288539
A;Accession: A36950
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-100 <MAE>
A;Cross-references: UNIPROT:Q07399; GB:D14439; NID:g393296; PIDN:BAA03323.1; PID:g21636
C;Superfamily: urease, gamma subunit; urease 11K chain homology
C;Keywords: hydrolase
F;1-100/Domain: urease 11K chain homology <U11>

  Query Match             64.8%;  Score 35;  DB 2;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 RKERGLK 11
              |||||||
Db         23 RKERGLK 29


RESULT 10
S49963
hypothetical protein YIL019w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YI3299.12
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C;Accession: S49963
R;Skelton, J.; Churcher, C.
submitted to the EMBL Data Library, December 1994
A;Reference number: S49951
A;Accession: S49963
A;Molecule type: DNA
A;Residues: 1-346 <SKE>
A;Cross-references: UNIPROT:P40546; EMBL:Z46881; NID:g599967; PIDN:CAA86973.1; PID:g599
C;Genetics:
A;Gene: MIPS:YIL019w
A;Cross-references: SGD:S0001281
A;Map position: 9L

  Query Match             64.8%;  Score 35;  DB 2;  Length 346;
  Best Local Similarity   60.0%;  Pred. No. 59;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGLK 11
              : |::|||||
Db        298 KATKRERGLK 307


RESULT 11
AC1031
```

```
hypothetical protein STY4573 [imported] - Salmonella enterica subsp. enterica serovar Ty
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC1031
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC1031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-807 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD09348.1; PID:g16505348; GSPDB:GN00176
C;Genetics:
A;Gene: STY4573

  Query Match             64.8%;  Score 35;  DB 2;  Length 807;
  Best Local Similarity   77.8%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGL 10
              |||| :|||
Db        508 RLTRADRGL 516


RESULT 12
C60950
apolipoprotein B-100 - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: C60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LDL
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: C60950
A;Molecule type: DNA
A;Residues: 1-269 <LAW>
A;Cross-references: UNIPROT:Q60537; UNIPROT:Q60536
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;

  Query Match             63.9%;  Score 34.5;  DB 2;  Length 269;
  Best Local Similarity   81.8%;  Pred. No. 59;
  Matches    9;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              :||||| ||||
Db        216 SRLTRK-RGLK 225


RESULT 13
JH0102
apolipoprotein B - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: JH0102
R;Smith, T.J.
submitted to GenBank, June 1990
A;Reference number: A38864
A;Accession: JH0102
A;Molecule type: DNA
A;Residues: 1-779 <SMI>
A;Cross-references: UNIPROT:Q60536; GB:M35187
A;Note: this is a revision to the sequence from reference JH0101
R;Smith, T.J.; Hautamaa, D.; Maeda, N.
Gene 87, 309-310, 1990
A;Title: Sequence of the putative low-density lipoprotein receptor-binding regions of ap
A;Reference number: JH0101; MUID:90236327; PMID:2332175
A;Contents: annotation
A;Note: this sequence has been revised in reference A38864
C;Genetics:
A;Gene: apoB
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotei
F;435-445/Region: receptor binding
F;646-656/Region: receptor binding

  Query Match             63.9%;  Score 34.5;  DB 2;  Length 779;
  Best Local Similarity   81.8%;  Pred. No. 1.5e+02;
  Matches    9;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              :||||| ||||
Db        642 SRLTRK-RGLK 651


RESULT 14
B64188
arginine transport system permease protein artQ - Haemophilus influenzae (strain Rd KW2
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C;Accession: B64188
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: B64188
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-221 <TIGR>
A;Cross-references: UNIPROT:P45090; GB:U32797; GB:L42023; NID:g1574095; PIDN:AAC22829.1
C;Genetics:
A;Gene: artQ
C;Function:
A;Description: probably required for translocation of the substrate across the membrane
C;Superfamily: histidine permease protein M
C;Keywords: amino acid transport; transmembrane protein
F;4-20/Domain: transmembrane #status predicted <TM1>
F;21-37/Domain: transmembrane #status predicted <TM2>
F;50-66/Domain: transmembrane #status predicted <TM3>
F;84-100/Domain: transmembrane #status predicted <TM4>
F;190-206/Domain: transmembrane #status predicted <TM5>

  Query Match             63.0%;  Score 34;  DB 2;  Length 221;
  Best Local Similarity   70.0%;  Pred. No. 61;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGLK 11
              | || |||:|
Db        212 RFTRFERGVK 221


RESULT 15
A69843
hypothetical protein yjbA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: A69843
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
```

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69843
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-250 <KUN>
A;Cross-references: UNIPROT:O31597; GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB12998.
A;Experimental source: strain 168
C;Genetics:
A;Gene: yjbA
C;Superfamily: Bacillus subtilis hypothetical protein yjbA

  Query Match 63.0%; Score 34; DB 2; Length 250;
  Best Local Similarity 77.8%; Pred. No. 69;
  Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          3 LTRKERGLK 11
              |||||| :|
Db        157 LTRKERQIK 165
```

Search completed: January 13, 2005, 01:52:42
Job time : 17.8689 secs

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:15:45 ; Search time 85.8361 Seconds
(without alignments)
73.735 Million cell updates/sec

Title: US-09-823-418-17
Perfect score: 54
Sequence: 1 TRLTRKERGLK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	77.8	2073	1	BIME_EMENI	P24686 emericella
2	38.5	71.3	414	2	Q7YQR5	Q7yqr5 aotus vocif
3	38.5	71.3	596	2	Q28473	Q28473 macaca fasc
4	38.5	71.3	3262	2	Q13788	Q13788 homo sapien
5	38.5	71.3	4563	1	APB_HUMAN	P04114 homo sapien
6	38.5	71.3	4563	2	Q7Z600	Q7z600 homo sapien
7	37	68.5	431	2	Q7U8V5	Q7u8v5 synechococc
8	37	68.5	488	2	Q9LXV0	Q9lxv0 arabidopsis
9	37	68.5	488	2	Q8VZF9	Q8vzf9 arabidopsis
10	36	66.7	94	2	Q89I78	Q89i78 bradyrhizob
11	36	66.7	248	2	Q6HM04	Q6hm04 bacillus th
12	36	66.7	248	2	Q73BX2	Q73bx2 bacillus ce
13	36	66.7	248	2	Q81GL8	Q81gl8 bacillus ce
14	36	66.7	248	2	Q81TS7	Q81ts7 bacillus an
15	36	66.7	248	2	Q9K8Y0	Q9k8y0 bacillus ha
16	36	66.7	248	2	AAS40225	Aas40225 bacillus
17	36	66.7	248	2	AAT30275	Aat30275 bacillus
18	36	66.7	250	2	Q8DNJ3	Q8dnj3 streptococc
19	36	66.7	252	2	Q8ERU3	Q8eru3 oceanobacil
20	36	66.7	264	2	Q8YU98	Q8yu98 anabaena sp
21	36	66.7	374	2	Q8KAK1	Q8kak1 chlorobium
22	36	66.7	378	2	Q7P868	Q7p868 fusobacteri
23	36	66.7	378	2	Q8RIN6	Q8rin6 fusobacteri
24	36	66.7	405	2	Q9X0D4	Q9x0d4 thermotoga
25	36	66.7	1095	1	AT9B_MOUSE	P98195 mus musculu
26	35	64.8	79	2	Q6ZE28	Q6ze28 oryza sativ
27	35	64.8	79	2	BAC83466	Bac83466 oryza sat
28	35	64.8	100	1	URE3_BACSB	Q07399 bacillus sp
29	35	64.8	100	2	Q733J4	Q733j4 bacillus ce
30	35	64.8	100	2	AAS42569	Aas42569 bacillus
31	35	64.8	117	2	Q82B17	Q82b17 streptomyce
32	35	64.8	210	2	Q7RYL3	Q7ryl3 neurospora
33	35	64.8	259	2	Q88JG3	Q88jg3 pseudomonas
34	35	64.8	281	2	Q72VN4	Q72vn4 leptospira
35	35	64.8	281	2	Q8F989	Q8f989 leptospira
36	35	64.8	281	2	AAS68890	Aas68890 leptospir
37	35	64.8	322	2	Q6CU76	Q6cu76 kluyveromyc
38	35	64.8	340	2	Q8GLA1	Q8gla1 streptococc
39	35	64.8	345	1	DDL_WOLSU	Q7ma71 wolinella s
40	35	64.8	346	1	YIB9_YEAST	P40546 saccharomyc
41	35	64.8	346	2	Q6Q5G5	Q6q5g5 saccharomyc
42	35	64.8	346	2	AAS56415	Aas56415 saccharom
43	35	64.8	401	2	Q7S6B9	Q7s6b9 neurospora
44	35	64.8	438	2	Q99JK2	Q99jk2 mus musculu
45	35	64.8	439	2	Q8RUL7	Q8rul7 oryza sativ

ALIGNMENTS

RESULT 1
BIME_EMENI
ID BIME_EMENI STANDARD; PRT; 2073 AA.
AC P24686;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Negative regulator of mitosis.
GN Name=BIME;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90375468; PubMed=1697851;
RA Engle D.B., Osmani S.A., Osmani A.H., Rosborough S., Xiang X.,
RA Morris N.R.;
RT "A negative regulator of mitosis in Aspergillus is a putative
RT membrane-spanning protein.";
RL J. Biol. Chem. 265:16132-16137(1990).
CC -!- FUNCTION: Negative regulator of mitosis in E.nidulans. This
CC protein is part of a regulatory pathway that includes the nimA
CC protein kinase. It is required to prevent premature entry into
CC mitosis. Mutations to this protein both cause cells to enter
CC mitosis and prevent them from leaving mitosis.
CC -!- SIMILARITY: Contains 4 PC repeats.
CC --

CC --
DR EMBL; M59705; AAA51478.1; -.
DR PIR; A37879; BWASBE.
DR InterPro; IPR002015; APC_proteasome.
DR Pfam; PF01851; PC_rep; 4.
KW Mitosis; Repeat; Transmembrane.
FT DOMAIN 342 353 Nuclear localization signal (Potential).
FT TRANSMEM 1623 1643 Potential.
FT TRANSMEM 1685 1703 Potential.
FT TRANSMEM 1746 1764 Potential.
SQ SEQUENCE 2073 AA; 229178 MW; 05E4E81EADDF51E4 CRC64;

Query Match 77.8%; Score 42; DB 1; Length 2073;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TRLTRKERGL 10
              |||||::|||
Db        832 TRLTRRKRGL 841
```

```
RESULT 2
Q7YQR5
ID   Q7YQR5      PRELIMINARY;      PRT;   414 AA.
AC   Q7YQR5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Aotus vociferans (Spix's owl monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX   NCBI_TaxID=57176;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548396; AAP97352.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     414    414
SQ   SEQUENCE   414 AA;  45955 MW;  EEFA8492157E1BDE CRC64;

  Query Match             71.3%;  Score 38.5;  DB 2;  Length 414;
  Best Local Similarity   90.9%;  Pred. No. 50;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db        258 TRLTRK-RGLK 267


RESULT 3
Q28473
ID   Q28473      PRELIMINARY;      PRT;   596 AA.
AC   Q28473;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Apolipoprotein B (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92075708; PubMed=1742325;
RA   Pape M.E., Castle C.K., Murray R.W., Funk G.M., Hunt C.E.,
RA   Marotti K.R., Melchior G.W.;
RT   "Apo B metabolism in the cynomolgus monkey: evidence for post-
RT   transcriptional regulation.";
RL   Biochim. Biophys. Acta 1086:326-334(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Murray R.;
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X15737; CAA33755.1; -.
DR   PIR; S32802; S32802.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     596    596
SQ   SEQUENCE   596 AA;  66757 MW;  B13BBA74E25C3120 CRC64;

  Query Match             71.3%;  Score 38.5;  DB 2;  Length 596;
  Best Local Similarity   90.9%;  Pred. No. 74;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db        226 TRLTRK-RGLK 235


RESULT 4
Q13788
ID   Q13788      PRELIMINARY;      PRT;  3262 AA.
AC   Q13788;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   APOB protein (Fragment).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87191999; PubMed=2883086;
RA   Carlsson P., Darnfors C., Olofsson S.O., Bjursell G.;
RT   "Analysis of the human apolipoprotein B gene; complete structure of
RT   the B-74 region.";
RL   Gene 49:29-51(1986).
DR   EMBL; M15421; AAA51758.1; -.
DR   PIR; A27850; LPHUB.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0005319; F:lipid transporter activity; NAS.
DR   GO; GO:0006869; P:lipid transport; NAS.
FT   NON_TER       1      1
SQ   SEQUENCE   3262 AA;  370140 MW;  56603BC0618DD40D CRC64;

  Query Match             71.3%;  Score 38.5;  DB 2;  Length 3262;
  Best Local Similarity   90.9%;  Pred. No. 4.8e+02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db       2084 TRLTRK-RGLK 2093


RESULT 5
APB_HUMAN
ID   APB_HUMAN      STANDARD;      PRT;  4563 AA.
AC   P04114; O00502; Q13787;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
DE   B-48 (Apo B-48)].
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87016385; PubMed=3763409;
RA   Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
RA   Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
RT   "Complete cDNA and derived protein sequence of human apolipoprotein B-
RT   100.";
RL   Nucleic Acids Res. 14:7501-7503(1986).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT GLU-4181.
RX   MEDLINE=88003974; PubMed=3652907;
RA   Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
RA   Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
RT   "DNA sequence of the human apolipoprotein B gene.";
RL   DNA 6:363-372(1987).
```

RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS ILE-98 AND GLU-4181.
RX MEDLINE=87008488; PubMed=3759943;
RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
RA Gotto A.M. Jr., Chan L.;
RT "The complete cDNA and amino acid sequence of human apolipoprotein B-
RT 100.";
RL J. Biol. Chem. 261:12918-12921(1986).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=87041416; PubMed=3464946;
RA Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
RA Lee N., Brewer H.B. Jr.;
RT "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
RT derived amino acid sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=87161758; PubMed=3030729;
RA Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
RA Zannis V.I.;
RT "The complete sequence and structural analysis of human apolipoprotein
RT B-100: relationship between apoB-100 and apoB-48 forms.";
RL EMBO J. 5:3495-3507(1986).
RN [6]
RP SEQUENCE OF 709-906 FROM N.A.
RX MEDLINE=85270450; PubMed=3860836;
RA Deeb S.S., Motulsky A.G., Albers J.J.;
RT "A partial cDNA clone for human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN [7]
RP SEQUENCE OF 3056-3159 FROM N.A.
RX MEDLINE=86041888; PubMed=3903660;
RA Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
RA Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
RT "Human apolipoprotein B: identification of cDNA clones and
RT characterization of mRNA.";
RL Nucleic Acids Res. 13:6937-6953(1985).
RN [8]
RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A., AND VARIANT GLU-4181.
RX MEDLINE=86093680; PubMed=3841204;
RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
RA Bjursell G.;
RT "Molecular cloning of human apolipoprotein B cDNA.";
RL Nucleic Acids Res. 13:8813-8826(1985).
RN [9]
RP SEQUENCE OF 3109-4563 FROM N.A.
RX MEDLINE=85300528; PubMed=2994225;
RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Urdea M.S.,
RA Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M.,
RA Priestley L.M., Robertson E., Rall L.B., Betsholtz C., Shows T.B.,
RA Mahley R.W., Scott J.;
RT "Human apolipoprotein B: structure of carboxyl-terminal domains, sites
RT of gene expression, and chromosomal localization.";
RL Science 230:37-43(1985).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
RT "Isolation of a cDNA clone encoding the amino-terminal region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A., AND VARIANT ILE-98.
RX MEDLINE=86287319; PubMed=3461454;
RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanaka M.,
RA Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT "Analysis of cDNA clones encoding the entire B-26 region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX MEDLINE=88018019; PubMed=3659919;
RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
RA Gotto A.M. Jr., Li W.-H., Chan L.;
RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA Levy-Wilson B., Scott J.;
RT "Complete protein sequence and identification of structural domains of
RT human apolipoprotein B.";
RL Nature 323:734-738(1986).
RN [14]
RP DOMAINS.
RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT "Sequence, structure, receptor-binding domains and internal repeats of
RT human apolipoprotein B-100.";
RL Nature 323:738-742(1986).
RN [15]
RP CALCIUM-BINDING DATA.
RX MEDLINE=86242245; PubMed=3087360;
RA Dashti N., Lee D.M., Mok T.;
RT "Apolipoprotein B is a calcium binding protein.";
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN [16]
RP PALMITOYLATION OF CYS-1112.
RX MEDLINE=20143590; PubMed=10679026;
RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT "Palmitoylation of apolipoprotein B is required for proper
RT intracellular sorting and transport of cholesteroyl esters and
RT triglycerides.";
RL Mol. Biol. Cell 11:721-734(2000).
RN [17]
RP VARIANT SER-4338.
RX MEDLINE=91071750; PubMed=1979313;
RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA Cunny G., Cambien F., Roizes G.;
RT "Detection by denaturing gradient gel electrophoresis of a new
RT polymorphism in the apolipoprotein B gene.";
RL Hum. Genet. 86:91-93(1990).
RN [18]
RP VARIANT FDB GLN-3527.
RX MEDLINE=89098975; PubMed=2563166;
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA McCarthy B.J.;
RT "Association between a specific apolipoprotein B mutation and familial
RT defective apolipoprotein B-100.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN [19]
RP VARIANT LEU-2739.
RX MEDLINE=91016974; PubMed=2216805;
RA Huang L.-S., Gavish D., Breslow J.L.;
RT "Sequence polymorphism in the human apoB gene at position 8344.";
RL Nucleic Acids Res. 18:5922-5922(1990).
RN [20]
RP VARIANT FDB CYS-3558.
RX MEDLINE=95190020; PubMed=7883971;
RA Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT "Familial ligand-defective apolipoprotein B. Identification of a new
RT mutation that decreases LDL receptor binding affinity.";
RL J. Clin. Invest. 95:1225-1234(1995).
RN [21]
RP VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP AND THR-4481.
RX MEDLINE=97044521; PubMed=8889592;

```
RA   Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA   Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT   "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT   PCR-SSCP.";
RL   Hum. Mutat. 8:282-285(1996).
RN   [22]
RP   VARIANTS FDB GLN-3527 AND CYS-3558.
RX   MEDLINE=97403938; PubMed=9259199;
RA   Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA   Krempf M., Giraudet P., Junien C., Boileau C.;
RT   "Familial ligand-defective apolipoprotein B-100: simultaneous
RT   detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a French
RT   population.";
RL   Hum. Mutat. 10:160-163(1997).
RN   [23]
RP   VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP   AND ILE-3921.
RX   MEDLINE=98141125; PubMed=9490296;
RA   Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT   "Screening for mutations of the apolipoprotein B gene causing
RT   hypocholesterolemia.";
RL   Hum. Genet. 102:44-49(1998).
CC   -!- FUNCTION: Apolipoprotein B is a major protein constituent of
CC       chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo
CC       B-100 functions as a recognition signal for the cellular binding
CC       and internalization of LDL particles by the apoB/E receptor.
CC   -!- SUBCELLULAR LOCATION: Secreted.

  Query Match             71.3%;  Score 38.5;  DB 1;  Length 4563;
  Best Local Similarity   90.9%;  Pred. No. 6.9e+02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db       3385 TRLTRK-RGLK 3394


RESULT 6
Q7Z600
ID   Q7Z600      PRELIMINARY;      PRT;  4563 AA.
AC   Q7Z600;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Apolipoprotein B (Including Ag(X) antigen).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA   Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY324608; AAP72970.1; -.
DR   GO; GO:0005319; F:lipid transporter activity; IEA.
DR   GO; GO:0006869; P:lipid transport; IEA.
DR   InterPro; IPR009454; DUF1081.
DR   InterPro; IPR001747; Lipid_transprt_N.
DR   Pfam; PF06448; DUF1081; 1.
DR   Pfam; PF01347; Vitellogenin_N; 1.
DR   SMART; SM00638; LPD_N; 1.
KW   Lipoprotein.
SQ   SEQUENCE   4563 AA;  515553 MW;  030B34167CEDC63C CRC64;

  Query Match             71.3%;  Score 38.5;  DB 2;  Length 4563;
  Best Local Similarity   90.9%;  Pred. No. 6.9e+02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKERGLK 11
              |||||| ||||
Db       3385 TRLTRK-RGLK 3394


RESULT 7
Q7U8V5
ID   Q7U8V5      PRELIMINARY;      PRT;  431 AA.
AC   Q7U8V5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Possible GTPase.
GN   OrderedLocusNames=SYNW0504;
OS   Synechococcus sp. (strain WH8102).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=84588;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22825697; PubMed=12917641; DOI=10.1038/nature01943;
RA   Palenik B., Brahamsha B., Larimer F.W., Land M.L., Hauser L.,
RA   Chain P., Lamerdin J.E., Regala W., Allen E.E., McCarren J.,
RA   Paulsen I.T., Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT   "The genome of a motile marine Synechococcus.";
RL   Nature 424:1037-1042(2003).
DR   EMBL; BX569690; CAE07019.1; -.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   InterPro; IPR005225; Small_GTP.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
KW   Complete proteome.
SQ   SEQUENCE   431 AA;  47647 MW;  666ECD647A83217D CRC64;

  Query Match             68.5%;  Score 37;  DB 2;  Length 431;
  Best Local Similarity   77.8%;  Pred. No. 1e+02;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGL 10
              ||||:|:||
Db         27 RLTRREQGL 35


RESULT 8
Q9LXV0
ID   Q9LXV0      PRELIMINARY;      PRT;  488 AA.
AC   Q9LXV0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Glucosyltransferase-like protein.
GN   Name=T24H18_60;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Robben J., Grymonprez B., Volckaert G., Bancroft I.,
RA   Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
DR   EMBL; AL353013; CAB88253.1; -.
DR   PIR; T49903; T49903.
DR   GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR002213; UDP_glucos_trans.
DR   Pfam; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   488 AA;  54867 MW;  B6CC2E0A55452647 CRC64;
```

```
  Query Match             68.5%;  Score 37;  DB 2;  Length 488;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGL 10
              |:|| ||||
Db        344 RITRSERGL 352


RESULT 9
Q8VZF9
ID   Q8VZF9      PRELIMINARY;      PRT;   488 AA.
AC   Q8VZF9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   AT5g12890/T24H18_60.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Meyers M.C., Banh J., Bowser L.,
RA   Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA   Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA   Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the UDP-glycosyltransferase family.
DR   EMBL; AY064985; AAL57638.1; -.
DR   GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   InterPro; IPR002213; UDP_glucos_trans.
DR   Pfam; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   488 AA;  54793 MW;  B6CC3E07FE389D37 CRC64;

  Query Match             68.5%;  Score 37;  DB 2;  Length 488;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKERGL 10
              |:|| ||||
Db        344 RITRSERGL 352


RESULT 10
Q89I78
ID   Q89I78      PRELIMINARY;      PRT;    94 AA.
AC   Q89I78;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Bsl5761 protein.
GN   OrderedLocusNames=bsl5761;
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005956; BAC51026.1; -.
KW   Complete proteome.
SQ   SEQUENCE   94 AA;  10344 MW;  4925ED8E76A04BB8 CRC64;

  Query Match             66.7%;  Score 36;  DB 2;  Length 94;
  Best Local Similarity   87.5%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKERG 9
              ||||:|||
Db         86 RLTREERG 93


RESULT 11
Q6HM04
ID   Q6HM04      PRELIMINARY;      PRT;   248 AA.
AC   Q6HM04;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=BT9727_1080;
OS   Bacillus thuringiensis serovar konkukian str. 97-27.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC   Bacillus thuringiensis serovar konkukian.
OX   NCBI_TaxID=281309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97-27;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017355; AAT63074.1; -.
DR   InterPro; IPR010983; EF_Hand_like.
KW   Hypothetical protein.
SQ   SEQUENCE   248 AA;  30052 MW;  4C44DCDD6B421736 CRC64;

  Query Match             66.7%;  Score 36;  DB 2;  Length 248;
  Best Local Similarity   88.9%;  Pred. No. 91;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 LTRKERGLK 11
              ||||| ||
Db        155 LTRKERQLK 163


RESULT 12
Q73BX2
ID   Q73BX2      PRELIMINARY;      PRT;   248 AA.
AC   Q73BX2;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=BCE1296;
OS   Bacillus cereus (strain ATCC 10987).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=222523;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14960714;
RA   Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA   Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA   Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT   "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT   adaptations and a large plasmid related to Bacillus anthracis pXO1.";
RL   Nucleic Acids Res. 32:977-988(2004).
DR   EMBL; AE017268; AAS40225.1; -.
DR   TIGR; BCE1296; -.
DR   InterPro; IPR010983; EF_Hand_like.
```

```
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   248 AA;  30064 MW;  2EC2A2C58C7CA2EA CRC64;

  Query Match             66.7%;  Score 36;  DB 2;  Length 248;
  Best Local Similarity   88.9%;  Pred. No. 91;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 LTRKERGLK 11
              |||||| ||
Db        155 LTRKERQLK 163


RESULT 13
Q81GL8
ID   Q81GL8      PRELIMINARY;      PRT;   248 AA.
AC   Q81GL8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=BC1176;
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.C.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017001; AAP08163.1; -.
DR   InterPro; IPR010983; EF_Hand_like.
KW   Hypothetical protein.
SQ   SEQUENCE   248 AA;  29788 MW;  79EBCC5746C2BDA3 CRC64;

  Query Match             66.7%;  Score 36;  DB 2;  Length 248;
  Best Local Similarity   88.9%;  Pred. No. 91;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 LTRKERGLK 11
              |||||| ||
Db        155 LTRKERQLK 163


RESULT 14
Q81TS7
ID   Q81TS7      PRELIMINARY;      PRT;   248 AA.
AC   Q81TS7; Q6I210; Q6KVU7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=BA1187, BAS1098; ORFNames=GBAA1187;
OS   Bacillus anthracis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1392;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames / isolate Porton;
RX   MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA   Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA   Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M.,
RA   Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames / isolate 0581;
RA   Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA   Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA   Fraser C.M.;
RT   "Bacillus anthracis comparative genomics.";
RL   Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sterne;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017027; AAP25150.1; -.
DR   EMBL; AE017334; AAT30275.1; -.
DR   EMBL; AE017225; AAT53421.1; -.
DR   TIGR; BA1187; -.
DR   InterPro; IPR010983; EF_Hand_like.
KW   Hypothetical protein.
SQ   SEQUENCE   248 AA;  30052 MW;  4C44DCDD6B421736 CRC64;

  Query Match             66.7%;  Score 36;  DB 2;  Length 248;
  Best Local Similarity   88.9%;  Pred. No. 91;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 LTRKERGLK 11
              |||||| ||
Db        155 LTRKERQLK 163


RESULT 15
Q9K8Y0
ID   Q9K8Y0      PRELIMINARY;      PRT;   248 AA.
AC   Q9K8Y0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   BH2872 protein.
GN   Name=BH2872;
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001516; BAB06591.1; -.
DR   PIR; H84008; H84008.
SQ   SEQUENCE   248 AA;  29874 MW;  C75106082942A7DE CRC64;

  Query Match             66.7%;  Score 36;  DB 2;  Length 248;
  Best Local Similarity   88.9%;  Pred. No. 91;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 LTRKERGLK 11
              |||||| ||
Db        155 LTRKERQLK 163
```

Search completed: January 13, 2005, 01:51:08
Job time : 88.8361 secs

THIS PAGE IS BLANK

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:14:40 ; Search time 84.9344 Seconds
(without alignments)
46.460 Million cell updates/sec

Title: US-09-823-418-19
Perfect score: 55
Sequence: 1 TRLTRKDRGLK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55	100.0	11	2	AAY30700	Aay30700 Apo-B100
2	51	92.7	11	2	AAY30698	Aay30698 Apo-B100
3	44	80.0	11	2	AAY30697	Aay30697 Apo-B100
4	42	76.4	11	2	AAY30699	Aay30699 Apo-B100
5	39.5	71.8	10	2	AAY30683	Aay30683 Apo-B100
6	39.5	71.8	10	2	AAY30691	Aay30691 Apo-B100
7	39	70.9	621	6	ABU24550	Abu24550 Protein e
8	39	70.9	1756	8	ADN47108	Adn47108 Thermococ
9	38.5	70.0	11	2	AAW57205	Aaw57205 Apo B bin
10	38.5	70.0	13	2	AAW57207	Aaw57207 Apo B 100
11	38.5	70.0	15	2	AAW41261	Aaw41261 Apolipopr
12	38.5	70.0	15	2	AAW96892	Aaw96892 ApoB-100
13	38.5	70.0	20	6	ABJ37575	Abj37575 Heparin b
14	38.5	70.0	22	2	AAW57208	Aaw57208 Apo B 100
15	38.5	70.0	22	2	AAW57209	Aaw57209 Apo B 100
16	38.5	70.0	34	5	AAE14541	Aae14541 Human apo
17	38.5	70.0	36	2	AAW96876	Aaw96876 Nucleic a
18	38.5	70.0	37	2	AAW64587	Aaw64587 Human apo
19	38.5	70.0	51	2	AAW96845	Aaw96845 Nucleic a
20	38.5	70.0	343	4	ABB37687	Abb37687 Peptide #
21	38.5	70.0	343	4	ABG52504	Abg52504 Human liv
22	38.5	70.0	377	2	AAR72704	Aar72704 Human apo
23	38.5	70.0	377	2	AAR34031	Aar34031 Sequence
24	38.5	70.0	2463	8	ADJ57400	Adj57400 Human apo
25	38.5	70.0	3923	2	AAY31237	Aay31237 Human Apo
26	38.5	70.0	4536	2	AAW41262	Aaw41262 Apolipopr
27	38.5	70.0	4536	2	AAW96826	Aaw96826 Amino aci
28	38.5	70.0	4560	5	AAU98981	Aau98981 Human apo
29	38.5	70.0	4561	7	ADD48677	Add48677 Human Pro
30	38.5	70.0	4563	5	AAO15893	Aao15893 Human apo
31	38.5	70.0	4563	6	ABR40253	Abr40253 Human ali
32	38.5	70.0	4563	6	ABU79140	Abu79140 Apolipopr
33	38.5	70.0	4563	7	ADF43408	Adf43408 Apolipopr
34	38.5	70.0	4563	8	ADH18871	Adh18871 Human apo
35	38.5	70.0	4563	8	ADH18870	Adh18870 Human apo
36	38.5	70.0	4563	8	ADO33445	Ado33445 Human apo
37	38.5	70.0	4563	8	ADO33447	Ado33447 Human apo
38	38.5	70.0	4590	4	AAU33184	Aau33184 Novel hum
39	38	69.1	97	4	AAU21930	Aau21930 Human car
40	38	69.1	97	7	ADE45898	Ade45898 Human car
41	37	67.3	203	4	AAE10805	Aae10805 Oryza sat
42	37	67.3	203	6	ADA89379	Ada89379 Rice hype
43	37	67.3	265	6	ADA89383	Ada89383 Soybean h
44	37	67.3	269	6	ADA89391	Ada89391 Rice hype
45	37	67.3	270	6	ADA89389	Ada89389 Rice hype

ALIGNMENTS

RESULT 1
AAY30700
ID AAY30700 standard; peptide; 11 AA.
XX
AC AAY30700;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 58; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can

CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 55; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKDRGLK 11
||||||||||||
Db 1 TRLTRKDRGLK 11

RESULT 2
AAY30698
ID AAY30698 standard; peptide; 11 AA.
XX
AC AAY30698;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 11 AA;

Query Match 92.7%; Score 51; DB 2; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0043;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKDRGLK 11
||||||:||||
Db 1 TRLTRKERGLK 11

RESULT 3
AAY30697
ID AAY30697 standard; peptide; 11 AA.
XX
AC AAY30697;
XX
DT 17-NOV-1999 (first entry)
XX
DE Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9946598-A1.
XX
PD 16-SEP-1999.
XX
PF 05-MAR-1999; 99WO-US004805.
XX
PR 10-MAR-1998; 98US-0077618P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Innerarity TL, Boren JOS;
XX
DR WPI; 1999-551509/46.
XX
PT Identifying compounds which affect binding of low density lipoprotein
PT with proteoglycan, used for, e.g. obtaining compounds for reducing
PT atherosclerosis.
XX
PS Claim 17; Page 57; 70pp; English.
XX
CC AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC receptor mutations. They were created to identify compounds which
CC modulate atherosclerosis. The peptides are derived from amino acids 3358
CC to 3367 of apoB100. The method comprises detecting compounds which affect
CC low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC can be used for identifying compounds which disrupt LDL-PG binding
CC without inhibiting LDL receptor binding. Such compounds can be used to
CC reduce or prevent the formation of atherosclerotic lesions and prevent
CC atherosclerosis. The transgenic non-human animals and mammals which
CC express human apo-B100 can be used as an in vivo model system for the
CC study of atherosclerosis, and in vivo assay methods for identifying
CC compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC also be used to identify compounds which result in an increase in
CC atherosclerotic regions. Thus the assays may be used to determine whether
CC a particular food or drug composition tends to stimulate or inhibit the
CC formation of atherosclerotic lesions. The polynucleotides can also be
CC used in gene therapy for preventing or reducing the severity of
CC atherosclerosis in an animal or mammal
XX
SQ Sequence 11 AA;

Query Match 80.0%; Score 44; DB 2; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.1;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDRGLK 11

```
              |||||: ||||
Db          1 TRLTREKRGLK 11


RESULT 4
AAY30699
ID   AAY30699 standard; peptide; 11 AA.
XX
AC   AAY30699;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW   Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW   low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946598-A1.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US004805.
XX
PR   10-MAR-1998;   98US-0077618P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Innerarity TL,  Boren JOS;
XX
DR   WPI; 1999-551509/46.
XX
PT   Identifying compounds which affect binding of low density lipoprotein
PT   with proteoglycan, used for, e.g. obtaining compounds for reducing
PT   atherosclerosis.
XX
PS   Claim 17; Page 58; 70pp; English.
XX
CC   AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC   receptor mutations. They were created to identify compounds which
CC   modulate atherosclerosis. The peptides are derived from amino acids 3358
CC   to 3367 of apoB100. The method comprises detecting compounds which affect
CC   low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC   can be used for identifying compounds which disrupt LDL-PG binding
CC   without inhibiting LDL receptor binding. Such compounds can be used to
CC   reduce or prevent the formation of atherosclerotic lesions and prevent
CC   atherosclerosis. The transgenic non-human animals and mammals which
CC   express human apo-B100 can be used as an in vivo model system for the
CC   study of atherosclerosis, and in vivo assay methods for identifying
CC   compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 11 AA;

  Query Match             76.4%;  Score 42;  DB 2;  Length 11;
  Best Local Similarity   81.8%;  Pred. No. 0.26;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDRGLK 11
              ||||| ||||
Db          1 TRLTRDKRGLK 11


RESULT 5
AAY30683
ID   AAY30683 standard; peptide; 10 AA.
XX
AC   AAY30683;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
KW   Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW   low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946598-A1.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US004805.
XX
PR   10-MAR-1998;   98US-0077618P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Innerarity TL,  Boren JOS;
XX
DR   WPI; 1999-551509/46.
XX
PT   Identifying compounds which affect binding of low density lipoprotein
PT   with proteoglycan, used for, e.g. obtaining compounds for reducing
PT   atherosclerosis.
XX
PS   Claim 17; Page 57; 70pp; English.
XX
CC   AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC   receptor mutations. They were created to identify compounds which
CC   modulate atherosclerosis. The peptides are derived from amino acids 3358
CC   to 3367 of apoB100. The method comprises detecting compounds which affect
CC   low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC   can be used for identifying compounds which disrupt LDL-PG binding
CC   without inhibiting LDL receptor binding. Such compounds can be used to
CC   reduce or prevent the formation of atherosclerotic lesions and prevent
CC   atherosclerosis. The transgenic non-human animals and mammals which
CC   express human apo-B100 can be used as an in vivo model system for the
CC   study of atherosclerosis, and in vivo assay methods for identifying
CC   compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 10 AA;

  Query Match             71.8%;  Score 39.5;  DB 2;  Length 10;
  Best Local Similarity   90.9%;  Pred. No. 0.73;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              ||||| |||||
Db          1 TRLTR-DRGLK 10


RESULT 6
AAY30691
ID   AAY30691 standard; peptide; 10 AA.
XX
AC   AAY30691;
XX
DT   17-NOV-1999  (first entry)
XX
DE   Apo-B100 derived peptide showing a proteoglycan receptor mutation.
XX
```

```
KW   Apo-B100; proteoglycan receptor mutation; atherosclerosis;
KW   low density lipoprotein; proteoglycan; LDL; atherosclerotic lesion.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9946598-A1.
XX
PD   16-SEP-1999.
XX
PF   05-MAR-1999;   99WO-US004805.
XX
PR   10-MAR-1998;   98US-0077618P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Innerarity TL,  Boren JOS;
XX
DR   WPI; 1999-551509/46.
XX
PT   Identifying compounds which affect binding of low density lipoprotein
PT   with proteoglycan, used for, e.g. obtaining compounds for reducing
PT   atherosclerosis.
XX
PS   Claim 17; Page 57; 70pp; English.
XX
CC   AAY30582-Y30700 represent apo-B100 derived peptides showing proteoglycan
CC   receptor mutations. They were created to identify compounds which
CC   modulate atherosclerosis. The peptides are derived from amino acids 3358
CC   to 3367 of apoB100. The method comprises detecting compounds which affect
CC   low density lipoprotein (LDL) binding with proteoglycan (PG). The method
CC   can be used for identifying compounds which disrupt LDL-PG binding
CC   without inhibiting LDL receptor binding. Such compounds can be used to
CC   reduce or prevent the formation of atherosclerotic lesions and prevent
CC   atherosclerosis. The transgenic non-human animals and mammals which
CC   express human apo-B100 can be used as an in vivo model system for the
CC   study of atherosclerosis, and in vivo assay methods for identifying
CC   compounds which modulate atherosclerosis and/or LDL-PG binding. They can
CC   also be used to identify compounds which result in an increase in
CC   atherosclerotic regions. Thus the assays may be used to determine whether
CC   a particular food or drug composition tends to stimulate or inhibit the
CC   formation of atherosclerotic lesions. The polynucleotides can also be
CC   used in gene therapy for preventing or reducing the severity of
CC   atherosclerosis in an animal or mammal
XX
SQ   Sequence 10 AA;

  Query Match             71.8%;  Score 39.5;  DB 2;  Length 10;
  Best Local Similarity   90.9%;  Pred. No. 0.73;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              ||||||| |||
Db          1 TRLTRKD-GLK 10


RESULT 7
ABU24550
ID   ABU24550 standard; protein; 621 AA.
XX
AC   ABU24550;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #10077.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Clostridium botulinum.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA28420.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 52474; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 621 AA;

  Query Match             70.9%;  Score 39;  DB 6;  Length 621;
  Best Local Similarity   77.8%;  Pred. No. 68;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              |||:||||:
Db        228 RLTKKDRGV 236


RESULT 8
ADN47108
ID   ADN47108 standard; protein; 1756 AA.
XX
AC   ADN47108;
XX
DT   01-JUL-2004  (first entry)
```

XX
DE Thermococcus kodakaraensis KOD1 protein sequence SeqID986.
XX
KW gene disruption; gene targeting; marker gene; transformation;
KW homologous recombination; hyperthermostable archaebacterium; KOD1;
KW gene structure; gene function; enzyme activity; medicine;
KW forensic science; food; drug inspection; molecular biology; immunology.
XX
OS Thermococcus kodakaraensis.
XX
PN WO2004022736-A1.
XX
PD 18-MAR-2004.
XX
PF 29-AUG-2003; 2003WO-IB003597.
XX
PR 30-AUG-2002; 2002JP-00319011.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Imanaka T, Atomi H;
XX
DR WPI; 2004-257583/24.
XX
PT Method for disrupting targeted gene in genome of organism particularly
PT thermostable bacterium and with genome chips for analysis, applicable in
PT studying gene structure and functions.
XX
PS Claim 9; SEQ ID NO 986; 598pp; Japanese.
XX
CC This invention relates to a novel method for targeting disruption of an
CC arbitrary gene in a genome of an organism which comprises providing the
CC whole sequential data of the genome of such organism, selecting at least
CC 1 arbitrary region in the sequence, providing a vector that contains a
CC sequence homologous with the selected region and a marker gene,
CC transformation, and homologous recombination. The genome is preferably
CC the genome of a hyperthermostable archaebacterium, particularly
CC Thermococcus kodakaraensis KOD1. The method is for targeting the
CC disruption of a gene in the genome of an organism, which is applicable in
CC studying gene structure and functions as well as enzyme activities of
CC encoded proteins and useful in medicine, forensic science, food or drug
CC inspection, molecular biology and immunology. With this method, the
CC disruption of a gene at an arbitrary position in a genome can be achieved
CC efficiently and reliably. The present sequence is that of a protein
CC encoded by the genome of Thermococcus kodakaraensis which was derived
CC using the method of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1756 AA;

```
  Query Match             70.9%;  Score 39;  DB 8;  Length 1756;
  Best Local Similarity   88.9%;  Pred. No. 2e+02;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              ||| |||||
Db       1183 RLTSKDRGL 1191
```

RESULT 9
AAW57205
ID AAW57205 standard; peptide; 11 AA.
XX
AC AAW57205;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B binding site peptide 2.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
PN WO9813385-A2.
XX
PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 12; Page 52; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B binding site
CC peptide which can be used as a component of a non-naturally occurring,
CC receptor-competent low density lipoprotein (LDL) particle of the present
CC invention. The LDL particle comprises at least 1 peptide component that
CC has at least 1 binding site for an apo B protein receptor and at least 1
CC lipophilic substituent. Also described in the invention are peptides
CC containing an apo B binding sequence with at least 70% identity with
CC sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their dimers. Non-
CC naturally occurring, receptor-competent LDL particles are useful as: (i)
CC drug-targeting vectors for delivering anticancer drugs to cancer cells
CC that express an apo B protein receptor, and (ii) additives for cell
CC culture media especially as growth supplements. Non-naturally occurring,
CC receptor-competent LDL particles do not require the complete apo B
CC sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 11 AA;

```
  Query Match             70.0%;  Score 38.5;  DB 2;  Length 11;
  Best Local Similarity   90.9%;  Pred. No. 1.3;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db          2 TRLTRK-RGLK 11
```

RESULT 10
AAW57207
ID AAW57207 standard; peptide; 13 AA.
XX
AC AAW57207;
XX
DT 03-AUG-1998 (first entry)
XX
DE Apo B 100 binding site peptide analogue peptide B.
XX
KW Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW growth supplement; non-natural lipid particle; low density lipoprotein;
KW LDL; receptor component; apo B100 receptor site.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "attached to retinoic acid"
XX
PN WO9813385-A2.
XX

PD 02-APR-1998.
XX
PF 25-SEP-1997; 97WO-GB002610.
XX
PR 27-SEP-1996; 96GB-00020153.
XX
PA (UYST) UNIV STRATHCLYDE.
XX
PI Halbert GW, Owens MD, Baillie G;
XX
DR WPI; 1998-230637/20.
XX
PT Non-natural lipid particle comprising peptide binding to apo B protein
PT receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT that express this receptor.
XX
PS Claim 13; Fig 7; 73pp; English.
XX
CC The present sequence represents a specifically claimed Apo B 100 binding
CC site peptide analogue which can be used as a component of a non-
CC naturally occurring, receptor-competent low density lipoprotein (LDL)
CC particle of the present invention. The LDL particle comprises at least 1
CC peptide component that has at least 1 binding site for an apo B protein
CC receptor and at least 1 lipophilic substituent. Also described in the
CC invention are peptides containing an apo B binding sequence with at least
CC 70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC dimers. Non-naturally occurring, receptor-competent LDL particles are
CC useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC cancer cells that express an apo B protein receptor, and (ii) additives
CC for cell culture media especially as growth supplements. Non-naturally
CC occurring, receptor-competent LDL particles do not require the complete
CC apo B sequence, which is large and tends to aggregate, to provide binding
CC affinity to an apo B protein receptor
XX
SQ Sequence 13 AA;

Query Match 70.0%; Score 38.5; DB 2; Length 13;
Best Local Similarity 90.9%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKDRGLK 11
|||||| ||||
Db 3 TRLTRK-RGLK 12

RESULT 11
AAW41261
ID AAW41261 standard; peptide; 15 AA.
XX
AC AAW41261;
XX
DT 19-MAY-1998 (first entry)
XX
DE Apolipoprotein B-100 fragment.
XX
KW Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
KW angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW prothrombinase complex.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9743311-A1.
XX
PD 20-NOV-1997.
XX
PF 09-MAY-1997; 97WO-GB001255.
XX
PR 09-MAY-1996; 96GB-00009702.
XX
PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI Bruckdorfer KR, Ettelaie C;
XX
DR WPI; 1998-008798/01.
XX
PT Peptide fragments of apo:lipo:protein B-100 with anticoagulant activity -
PT used for treating or preventing coagulation, inhibiting angiogenesis,
PT cell differentiation and apoptosis.
XX
PS Disclosure; Page 22; 60pp; English.
XX
CC This sequence is an example of the peptide of the invention. It has the
CC formula (I), or their variants with one or more internal deletions,
CC insertions or substitutions, while retaining anti-coagulant properties of
CC apolipoprotein B-100 (apoB-100). Z1-KAQ-X1-KKNKHRHS-X2-T-Z2 (I) X1 = S or
CC Y; X2 = T or I; Z1 = the N terminus of the peptide, or 1-47 amino acids
CC (aa); Z2 = the C terminus of the peptide, a terminal amide group or 1-77
CC aa. Compositions containing the peptide are used for simultaneous,
CC separate or sequential treatment of cancer, particularly to prevent
CC metastatic spread. They are also used to inhibit thromboplastin-mediated
CC processes, specifically to prevent or reduce blood coagulation (e.g.
CC during or after surgery or in cases of heart attack, stroke etc.) and to
CC inhibit angiogenesis, cellular differentiation or apoptosis. KRAD-14,
CC which is active as such or as part of a 98-aa peptide, inhibits
CC activation of the prothrombinase complex; and prevents activation of
CC factor VII on the surface of thromboplastin and of platelets by thrombin.
CC It binds to the residues 58-66 of thromboplastin. Since (I) are much
CC smaller than apoB-100, they act more quickly
XX
SQ Sequence 15 AA;

Query Match 70.0%; Score 38.5; DB 2; Length 15;
Best Local Similarity 90.9%; Pred. No. 1.7;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKDRGLK 11
|||||| ||||
Db 1 TRLTRK-RGLK 10

RESULT 12
AAW96892
ID AAW96892 standard; peptide; 15 AA.
XX
AC AAW96892;
XX
DT 22-APR-1999 (first entry)
XX
DE ApoB-100 nuclear localisation signal sequence, residues 3353-3367.
XX
KW Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
XX
OS Homo sapiens.
XX
PN WO9856938-A1.
XX
PD 17-DEC-1998.
XX
PF 10-JUN-1998; 98WO-US011927.
XX
PR 13-JUN-1997; 97US-00874807.
PR 14-MAY-1998; 98US-00079030.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Guevara JG, Hoogeveen RC, Moore JP;
XX
DR WPI; 1999-070331/06.
XX
PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein -
PT used for delivering nucleic acid to cells for gene therapy and antisense

```
PT   treatment.
XX
PS   Claim 19; Fig 13D; 293pp; English.
XX
CC   AAW96878-97 represent nuclear localisation signal sequence derived from
CC   human apolipoprotein B-100 (apoB-100). ApoB-100 is a major apoprotein
CC   component of very-low density lipoproteins (VLDL), intermediate density
CC   lipoprotein (IDL), low density lipoproteins ((LDL) and lipoprotein a. The
CC   present sequence can be used in the composition of the invention. The
CC   specification describes a composition that comprises LDL and
CC   apolipoproteins for the binding and in vivo transport of nucleic acids.
CC   The composition is used to deliver nucleic acids to eukaryotic cells, in
CC   vivo or in vitro, for expressing a therapeutic polypeptide or antisense
CC   molecule (or ribozyme). Specifically they are used for gene therapy of
CC   cancers (particularly non-small cell lung carcinoma), diabetes, cystic
CC   fibrosis and arteriosclerosis
XX
SQ   Sequence 15 AA;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 15;
  Best Local Similarity   90.9%;  Pred. No. 1.7;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db          6 TRLTRK-RGLK 15


RESULT 13
ABJ37575
ID   ABJ37575 standard; peptide; 20 AA.
XX
AC   ABJ37575;
XX
DT   10-MAY-2003  (first entry)
XX
DE   Heparin binding peptide sequence #28.
XX
KW   Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW   cardiovascular; circulatory; ligand; sulphated; sulphonated; tumour;
KW   rheumatoid arthritis; hypoxia; diabetic retinopathy; heparin.
XX
OS   Unidentified.
XX
PN   WO2003007689-A2.
XX
PD   30-JAN-2003.
XX
PF   22-JUL-2002; 2002WO-US023419.
XX
PR   20-JUL-2001; 2001US-0306726P.
XX
PA   (ETHZ-) ETH ZUERICH.
PA   (UYZU-) UNIV ZURICH.
XX
PI   Hubbell JA,  Schoenmakers R,  Maynard HD;
XX
DR   WPI; 2003-300420/29.
XX
PT   Use of a ligand comprising of at least one sulfated or sulfonated amino
PT   acid for the treatment of e.g. tumors, rheumatoid arthritis, diabetic
PT   retinopathy and hypoxia.
XX
PS   Disclosure; Fig 2; 79pp; English.
XX
CC   The invention relates to a novel ligand for binding a target biomolecule,
CC   which comprises a peptide having at least one sulphated or sulphonated
CC   amino acid and at least one amino acid chosen from neutral and positively
CC   charged amino acids. The novel ligands can be used for the treatment of
CC   e.g. tumours, rheumatoid arthritis, diabetic retinopathy and hypoxia.
CC   This sequence represents a heparin binding peptide relating to the
CC   invention
XX
SQ   Sequence 20 AA;

  Query Match             70.0%;  Score 38.5;  DB 6;  Length 20;
  Best Local Similarity   90.9%;  Pred. No. 2.4;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db          7 TRLTRK-RGLK 16


RESULT 14
AAW57208
ID   AAW57208 standard; peptide; 22 AA.
XX
AC   AAW57208;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide C.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
FT   Modified-site   22
FT                   /note= "attached to cholesterol"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 22 AA;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 22;
  Best Local Similarity   90.9%;  Pred. No. 2.6;
```

```
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db          7 TRLTRK-RGLK 16


RESULT 15
AAW57209
ID   AAW57209 standard; peptide; 22 AA.
XX
AC   AAW57209;
XX
DT   03-AUG-1998  (first entry)
XX
DE   Apo B 100 binding site peptide analogue peptide D.
XX
KW   Apo B; binding site; receptor; cancer; drug delivery; anticancer;
KW   growth supplement; non-natural lipid particle; low density lipoprotein;
KW   LDL; receptor component; apo B100 receptor site.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "attached to retinoic acid"
XX
PN   WO9813385-A2.
XX
PD   02-APR-1998.
XX
PF   25-SEP-1997;   97WO-GB002610.
XX
PR   27-SEP-1996;   96GB-00020153.
XX
PA   (UYST ) UNIV STRATHCLYDE.
XX
PI   Halbert GW,  Owens MD,  Baillie G;
XX
DR   WPI; 1998-230637/20.
XX
PT   Non-natural lipid particle comprising peptide binding to apo B protein
PT   receptor - useful as, e.g. vector for delivering drugs to cancer cells
PT   that express this receptor.
XX
PS   Claim 13; Fig 7; 73pp; English.
XX
CC   The present sequence represents a specifically claimed Apo B 100 binding
CC   site peptide analogue which can be used as a component of a non-
CC   naturally occurring, receptor-competent low density lipoprotein (LDL)
CC   particle of the present invention. The LDL particle comprises at least 1
CC   peptide component that has at least 1 binding site for an apo B protein
CC   receptor and at least 1 lipophilic substituent. Also described in the
CC   invention are peptides containing an apo B binding sequence with at least
CC   70% identity with sequences: KAEYKKNKHRH (1) or TTRLTRKRGLK (2), or their
CC   dimers. Non-naturally occurring, receptor-competent LDL particles are
CC   useful as: (i) drug-targeting vectors for delivering anticancer drugs to
CC   cancer cells that express an apo B protein receptor, and (ii) additives
CC   for cell culture media especially as growth supplements. Non-naturally
CC   occurring, receptor-competent LDL particles do not require the complete
CC   apo B sequence, which is large and tends to aggregate, to provide binding
CC   affinity to an apo B protein receptor
XX
SQ   Sequence 22 AA;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 22;
  Best Local Similarity   90.9%;  Pred. No. 2.6;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db          7 TRLTRK-RGLK 16


Search completed: January 13, 2005, 01:43:05
Job time : 86.1011 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:30:05 ; Search time 15.8689 Seconds
(without alignments)
66.696 Million cell updates/sec

Title: US-09-823-418-19
Perfect score: 55
Sequence: 1 TRLTRKDRGLK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	72.7	250	2	D98082	hypothetical prote
2	40	72.7	2073	1	BWASBE	bimE protein - Eme
3	39	70.9	807	2	AC1031	hypothetical prote
4	38.5	70.0	596	2	S32802	apolipoprotein B -
5	38.5	70.0	4563	1	LPHUB	apolipoprotein B-1
6	36	65.5	250	2	F95218	hypothetical prote
7	35	63.6	78	2	E81098	hypothetical prote
8	35	63.6	236	2	AI2432	hypothetical prote
9	35	63.6	321	2	A37842	hypothetical prote
10	35	63.6	384	2	AE1987	hypothetical prote
11	34.5	62.7	269	2	C60950	apolipoprotein B-1
12	34.5	62.7	779	2	JH0102	apolipoprotein B -
13	34	61.8	195	2	F71164	hypothetical prote
14	34	61.8	331	2	S73019	daunorubicin resis
15	34	61.8	389	2	D64337	16S rRNA 5'-region
16	34	61.8	405	2	C72305	transposase, IS605
17	34	61.8	419	2	A81651	conserved hypothet
18	34	61.8	458	2	E91092	probable invasion
19	34	61.8	458	2	A85938	probable invasion
20	34	61.8	488	2	T49903	glucosyltransferas
21	34	61.8	543	2	F96624	hypothetical prote
22	33	60.0	246	2	C97177	C-terminal domain
23	33	60.0	248	2	H84008	hypothetical prote
24	33	60.0	264	2	AI2112	hypothetical prote
25	33	60.0	289	2	T01531	hypothetical prote
26	33	60.0	305	1	NKVLHH	core antigen - her
27	33	60.0	341	2	AC1148	hypothetical prote
28	33	60.0	389	2	AD0722	glutamate dehydrog
29	33	60.0	419	2	G84799	hypothetical prote
30	33	60.0	429	2	H71491	hypothetical prote
31	33	60.0	503	2	H86146	hypothetical prote
32	33	60.0	826	2	E96720	probable alpha-amy
33	33	60.0	1451	2	I40325	dermonecrotic toxi
34	33	60.0	3187	2	JC5837	364K Golgi complex
35	32.5	59.1	275	2	E60950	apolipoprotein B-1
36	32	58.2	100	2	A36950	urease (EC 3.5.1.5
37	32	58.2	107	2	B85356	glycine-rich prote
38	32	58.2	196	2	S69732	hypothetical prote
39	32	58.2	290	2	H72071	ct386 hypothetical
40	32	58.2	290	2	E86551	CT386 hypothetical
41	32	58.2	299	2	D83010	probable binding p
42	32	58.2	300	2	A81597	conserved hypothet
43	32	58.2	326	2	JQ0855	hypothetical 36.8K
44	32	58.2	346	2	S49963	hypothetical prote
45	32	58.2	398	2	S40752	hypothetical prote

ALIGNMENTS

RESULT 1
D98082
hypothetical protein fecE [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: D98082
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: D98082
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <KUR>
A;Cross-references: UNIPROT:Q8DNJ3; GB:AE007317; PIDN:AAL00489.1; PID:g15459361; GSPDB
C;Genetics:
A;Gene: fecE
C;Superfamily: inner membrane protein malK; ATP-binding cassette homology

Query Match 72.7%; Score 40; DB 2; Length 250;
Best Local Similarity 70.0%; Pred. No. 2.8;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TRLTRKDRGL 10
              :|||:||||:
Db         47 SRLTKKDRGV 56
```

RESULT 2
BWASBE
bimE protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C;Accession: A37879
R;Engle, D.B.; Osmani, S.A.; Osmani, A.H.; Rosborough, S.; Xiang, X.; Morris, N.R.
J. Biol. Chem. 265, 16132-16137, 1990
A;Title: A negative regulator of mitosis in Aspergillus is a putative membrane-spanning
A;Reference number: A37879; MUID:90375468; PMID:1697851
A;Accession: A37879
A;Molecule type: mRNA
A;Residues: 1-2073 <ENG>
A;Cross-references: UNIPROT:P24686; GB:M59705; GB:J05607; NID:g168026; PIDN:AAA51478.1,
A;Note: in addition to three predicted transmembrane domains, there are several potenti
asein kinase, and one sequence that resembles a nuclear localization signal
C;Comment: This protein is part of a regulatory pathway that includes the nimA protein
ter mitosis and prevent them from leaving mitosis.
C;Genetics:
A;Gene: bimE
C;Superfamily: bimE protein

```
C;Keywords: cell cycle control; mitosis; transmembrane protein
F;1623-1643/Domain: transmembrane #status predicted <TM1>
F;1685-1703/Domain: transmembrane #status predicted <TM2>
F;1746-1764/Domain: transmembrane #status predicted <TM3>

  Query Match             72.7%;  Score 40;  DB 1;  Length 2073;
  Best Local Similarity   80.0%;  Pred. No. 23;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDRGL 10
              |||||: |||
Db        832 TRLTRRKRGL 841


RESULT 3
AC1031
hypothetical protein STY4573 [imported] - Salmonella enterica subsp. enterica serovar Ty
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC1031
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC1031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-807 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD09348.1; PID:g16505348; GSPDB:GN00176
C;Genetics:
A;Gene: STY4573

  Query Match             70.9%;  Score 39;  DB 2;  Length 807;
  Best Local Similarity   88.9%;  Pred. No. 14;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              |||| ||||
Db        508 RLTRADRGL 516


RESULT 4
S32802
apolipoprotein B - crab-eating macaque (fragment)
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S32802
R;Pape, M.E.; Castle, C.K.; Murray, R.W.; Funk, G.M.; Hunt, C.E.; Marotti, K.R.; Melchio
Biochim. Biophys. Acta 1086, 326-334, 1991
A;Title: Apo B metabolism in the cynomolgus monkey: evidence for post-transcriptional re
A;Reference number: S32802; MUID:92075708; PMID:1742325
A;Accession: S32802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-596 <PAP>
A;Cross-references: UNIPROT:Q28473; EMBL:X15737; NID:g38047; PIDN:CAA33755.1; PID:g93012
C;Superfamily: apolipoprotein B

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 596;
  Best Local Similarity   90.9%;  Pred. No. 13;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db        226 TRLTRK-RGLK 235


RESULT 5
LPHUB
apolipoprotein B-100 precursor - human
N;Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
C;Species: Homo sapiens (man)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 09-Jul-2004
C;Accession: A27850; A25679; A25263; A25267; A25266; A24320; A24684; A23817; A25774; A2
4452; I61909; I59510; I39474; I39469; I84624; I37179; PS0058
R;Ludwig, E.H.; Blackhart, B.D.; Pierotti, V.R.; Caiati, L.; Fortier, C.; Knott, T.; Sc
DNA 6, 363-372, 1987
A;Title: DNA sequence of the human apolipoprotein B gene.
A;Reference number: A27850; MUID:88003974; PMID:3652907
A;Accession: A27850
A;Molecule type: DNA
A;Residues: 1-617,'A',619-1929,'F',1931-3318,'D',3320-3426,'T',3428-3431,'Q',3433-3731,
A;Cross-references: UNIPROT:P04114; UNIPROT:P78482; UNIPROT:P78479; UNIPROT:Q9UMN0; UNI
R;Cladaras, C.; Hadzopoulou-Cladaras, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A;Title: The complete sequence and structural analysis of human apolipoprotein B-100: r
A;Reference number: A91058; MUID:87161758; PMID:3030729
A;Accession: A25679
A;Molecule type: mRNA
A;Residues: 1-11,15-2539,'S',2541-3823,'R',3825-4563 <CLA>
A;Note: 1109-Asp was also found
R;Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McC
Nucleic Acids Res. 14, 7501-7503, 1986
A;Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A;Reference number: A93639; MUID:87016385; PMID:3763409
A;Accession: A25263
A;Molecule type: mRNA
A;Residues: 1-272,'N',274-617,'A',619-1217,'E',1219-2091,'V',2093-2364,'T',2366-2679,'Q
A;Cross-references: GB:X04506; NID:g34330; PIDN:CAA28191.1; PID:g34331
R;Law, S.W.; Grant, S.M.; Higuchi, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer J
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A;Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A;Reference number: A94134; MUID:87041416; PMID:3464946
A;Accession: A25267
A;Molecule type: mRNA
A;Residues: 1-617,'A',619-703,'P',705-792,'R',794-1270,'S',1272-1866,'G',1868-2036,'N',
4189-4220,'M',4222-4563 <LAW>
A;Note: the codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4188, and
R;Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Gotto Jr., A.
J. Biol. Chem. 261, 12918-12921, 1986
A;Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A;Reference number: A92556; MUID:87008488; PMID:3759943
A;Accession: A25266
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-328,'V',330-644,'I',646-918,'P',920-3318,'D',3320-3426,'T',3428
9-4132,'G',4134-4180,'E',4182-4563 <CHE>
A;Cross-references: GB:J02610; NID:g178803; PIDN:AAA35549.1; PID:g178804
A;Note: a total of 2366 residues were confirmed by direct sequencing of tryptic peptide
R;Protter, A.A.; Hardman, D.A.; Sato, K.Y.; Schilling, J.W.; Yamanaka, M.; Hort, Y.J.;
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A;Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotei
A;Reference number: A24320; MUID:86287319; PMID:3461454
A;Accession: A24320
A;Molecule type: mRNA
A;Residues: 1-97,'I',99-617,'A',619-941,'YYIWSLPPKP',951-1138,'PTGRLPNCFSNGLICYSLWLHSFQ
A;Cross-references: GB:M14081; NID:g178795; PIDN:AAA51752.1; PID:g553189
R;Law, S.W.; Lackner, K.J.; Hospattankar, A.V.; Anchors, J.M.; Sakaguchi, A.Y.; Naylor,
Proc. Natl. Acad. Sci. U.S.A. 82, 8340-8344, 1985
A;Title: Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of
A;Reference number: A24684; MUID:86094221; PMID:3001697
A;Accession: A24684
A;Molecule type: mRNA
A;Residues: 485-617,'A',619-1044 <LA2>
A;Cross-references: GB:M12480; NID:g178791; PIDN:AAA51751.1; PID:g178792
R;Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; K
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A;Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipop
A;Reference number: A94088; MUID:86149325; PMID:3513177
A;Accession: A23817
A;Molecule type: mRNA
A;Residues: 1-291 <PRO>
A;Cross-references: GB:M12681; NID:g178797; PIDN:AAA51753.1; PID:g178798
```

R;Deeb, S.S.; Motulsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A;Title: A partial cDNA clone for human apoliprotein B.
A;Reference number: A25774; MUID:85270450; PMID:3860836
A;Accession: A25774
A;Molecule type: mRNA
A;Residues: 709-791,'SSSWKAASHGCPHSAGD',810-906 <DEE>
A;Cross-references: GB:K03175; NID:g178821; PIDN:AAA51759.1; PID:g178822
R;Carlsson, P.; Darnfors, C.; Olofsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A;Title: Analysis of the human apolipoprotein B gene; complete structure of the B-74 reg
A;Reference number: A91565; MUID:87191999; PMID:2883086
A;Accession: A26533
A;Molecule type: mRNA
A;Residues: 1282-2721,2742-3290,'L',3292-3336,'N',3338-3948,'F',3950-3963,'Y',3965-4180,
A;Cross-references: GB:M15421; NID:g178817; PIDN:AAA51758.1; PID:g178818
R;Hardman, D.A.; Protter, A.A.; Chen, G.C.; Schilling, J.W.; Sato, K.Y.; Lau, K.; Yamana
Biochemistry 26, 5478-5486, 1987
A;Title: Structural comparison of human apolipoproteins B-48 and B-100.
A;Reference number: A29671; MUID:88050832; PMID:3676265
A;Accession: A29671
A;Molecule type: mRNA
A;Residues: 1671-2323,'PYW',2327-2352,'H',2354-2398 <HAR>
A;Cross-references: GB:M17367; NID:g178731; PIDN:AAA51741.1; PID:g178732
R;Shoulders, C.C.; Myant, N.B.; Sidoli, A.; Rodriguez, J.C.; Cortese, C.; Baralle, F.E.;
Atherosclerosis 58, 277-289, 1985
A;Title: Molecular cloning of human LDL apolipoprotein B cDNA. Evidence for more than or
A;Reference number: A90084; MUID:86130855; PMID:3841481
A;Accession: A29287
A;Molecule type: mRNA
A;Residues: 3846-4298 <SHO>
R;Pfitzner, R.; Wagener, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 367, 1077-1083, 1986
A;Title: Isolation, expression and characterization of a human apolipoprotein B 100-spec
A;Reference number: A25572; MUID:87076044; PMID:3024665
A;Accession: A25572
A;Molecule type: mRNA
A;Residues: 4219-4337,'S',4339-4563 <PFI>
A;Cross-references: GB:M36676
R;Wei, C.F.; Chen, S.H.; Yang, C.Y.; Marcel, Y.L.; Milne, R.W.; Li, W.H.; Sparrow, J.T.;
Proc. Natl. Acad. Sci. U.S.A. 82, 7265-7269, 1985
A;Reference number: A24738; MUID:86042646; PMID:2932736
A;Accession: A24738
A;Molecule type: mRNA
A;Residues: 'N',3729-3731,'I',3733-3875,'A',3877-3948,'F',3950-3963,'Y',3965-3982,'S',39
A;Cross-references: GB:M12413; NID:g178735; PIDN:AAA51742.1; PID:g178736
R;Chen, S.H.; Habib, G.; Yang, C.Y.; Gu, Z.W.; Lee, B.R.; Weng, S.; Silberman, S.R.; Cai
Science 238, 363-366, 1987
A;Title: Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in
A;Reference number: A40133; MUID:88018019; PMID:3659919
A;Accession: B40133
A;Molecule type: mRNA
A;Residues: 2165-2179 <CH1>
A;Cross-references: GB:M18036; NID:g178799; PIDN:AAA51754.1; PID:g178800
A;Note: this mRNA includes the stop codon of the organ-specific mRNA for apo48
A;Accession: A40133
A;Molecule type: protein
A;Residues: 51-75;101-110;129-139;158-174;197-207;276-287;298-304;306-314;526-532;538-55
36;1486-1498;1537-1556;1563-1572;1601-1610;1647-1661;1697-1724;1770-1781;1859-1897;1968-
A;Note: these fragments were derived from apo48
R;Hardman, D.A.; Protter, A.A.; Schilling, J.W.; Kane, J.P.
Biochem. Biophys. Res. Commun. 149, 1214-1219, 1987
A;Title: Carboxyl terminal analysis of human B-48 protein confirms the novel mechanism p
A;Reference number: A28002; MUID:88106542; PMID:3426612
A;Accession: A28002
A;Molecule type: mRNA
A;Residues: 2129-2179,2181-2235 <HA2>
A;Cross-references: GB:M18471
A;Experimental source: intestine
A;Note: this mRNA from intestine includes a stop codon created by RNA editing in place c
R;Mehrabian, M.; Schumaker, V.N.; Fareed, G.C.; West, R.; Johnson, D.F.; Kirchgessner, T
Nucleic Acids Res. 13, 6937-6953, 1985
A;Title: Human apolipoprotein B: identification of cDNA clones and characterization of m

A;Reference number: A24269; MUID:86041888; PMID:3903660
A;Accession: A24269
A;Molecule type: mRNA
A;Residues: 3056-3159 <MEH>
A;Cross-references: GB:X03045; NID:g28783; PIDN:CAA26850.1; PID:g929609
R;Hospattankar, A.V.; Higuchi, K.; Law, S.W.; Meglin, N.; Brewer Jr., H.B.
Biochem. Biophys. Res. Commun. 148, 279-285, 1987
A;Title: Identification of a novel in-frame translational stop codon in human intestine
A;Reference number: A29659; MUID:88049670; PMID:2445342
A;Accession: A29659
A;Molecule type: mRNA
A;Residues: 2169-2179 <HOS>
A;Note: the sequence shown represents the carboxyl end of apolipoprotein B-48
A;Note: two RNA species, 14.1kb and 7.5kb in length, were isolated from the human intes
ch encodes the 250K apoB-48, CAA encoding 2180-Gln is substituted by the stop codon TA
R;Yang, C.; Kim, T.W.; Weng, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
A;Title: Isolation and characterization of sulfhydryl and disulfide peptides of human a
A;Reference number: A35783; MUID:90319144; PMID:2115173
A;Contents: disulfide bonds
A;Accession: A35783
A;Molecule type: protein
A;Residues: 28-41;76-97,'I',99-100;175-193;206-215;239-249;259-266;357-399;455-490;512-

A;Note: cysteines at positions 1112, 1422, 1505, 1662, 3761, 3917, and 4217 have free s
R;LeBoeuf, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
FEBS Lett. 170, 105-108, 1984
A;Title: Human apolipoprotein B: partial amino acid sequence.
A;Reference number: A22006; MUID:84208786; PMID:6373369
A;Accession: A22006
A;Molecule type: protein
A;Residues: 873-892,'K',894-896 <LE1>
A;Accession: B22006
A;Molecule type: protein
A;Residues: 3113,'L',3115-3130,'R',3132-3133,'P',3135-3136,'R' <LE2>
R;Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Caiati, L.; Onasch, M.A.; Wallis, S.C
J. Biol. Chem. 261, 15364-15367, 1986
A;Title: Structure of the human apolipoprotein B gene.
A;Reference number: A92564; MUID:87057153; PMID:2946672
A;Contents: annotation; gene structure
R;Wagener, R.; Pfitzner, R.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
A;Title: Studies on the organization of the human apolipoprotein B 100 gene.
A;Reference number: A90715; MUID:87271140; PMID:2886136
A;Contents: annotation; gene structure
R;Weisgraber, K.H.; Rall Jr., S.C.
J. Biol. Chem. 262, 11097-11103, 1987
A;Title: Human apolipoprotein B-100 heparin-binding sites.
A;Reference number: A92605; MUID:87280197; PMID:3301850
A;Contents: annotation; heparin binding and disulfide bond
R;Dashti, N.; Lee, D.M.; Mok, T.
Biochem. Biophys. Res. Commun. 137, 493-499, 1986
A;Title: Apolipoprotein B is a calcium binding protein.
A;Reference number: A90125; MUID:86242245; PMID:3087360
A;Contents: annotation; calcium binding
R;Carlsson, P.; Olofsson, S.O.; Bondjers, G.; Darnfors, C.; Wiklund, O.; Bjursell, G.
Nucleic Acids Res. 13, 8813-8826, 1985
A;Title: Molecular cloning of human apolipoprotein B cDNA.
A;Reference number: I37178; MUID:86093680; PMID:3841204
A;Accession: I37180

  Query Match 70.0%; Score 38.5; DB 1; Length 4563;
  Best Local Similarity 90.9%; Pred. No. 98;
  Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TRLTRKDRGLK 11
  |||||| ||||
Db 3385 TRLTRK-RGLK 3394

RESULT 6
F95218
hypothetical protein SP1871 [imported] - Streptococcus pneumoniae (strain TIGR4)

C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: F95218
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: F95218
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <KUR>
A;Cross-references: UNIPROT:Q97NY0; GB:AE005672; PIDN:AAK75943.1; PID:g14973374; GSPDB:G
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1871
C;Superfamily: inner membrane protein malK; ATP-binding cassette homology

  Query Match 65.5%; Score 36; DB 2; Length 250;
  Best Local Similarity 60.0%; Pred. No. 18;
  Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TRLTRKDRGL 10
:|||:||:|:
Db 47 SRLTKKDQGV 56

RESULT 7
E81098
hypothetical protein NMB1311 [imported] - Neisseria meningitidis (strain MC58 serogroup
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: E81098
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: E81098
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-78 <TET>
A;Cross-references: UNIPROT:Q9JZ39; GB:AE002479; GB:AE002098; NID:g7226543; PIDN:AAF4168
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1311

  Query Match 63.6%; Score 35; DB 2; Length 78;
  Best Local Similarity 70.0%; Pred. No. 9;
  Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 RLTRKDRGLK 11
|| |||:| |
Db 26 RLLRKDKGRK 35

RESULT 8
AI2432
hypothetical protein all5017 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AI2432
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-236 <KUR>
A;Cross-references: UNIPROT:Q8YMB7; GB:BA000019; PIDN:BAB76716.1; PID:g17134155; GSPDB:
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all5017

  Query Match 63.6%; Score 35; DB 2; Length 236;
  Best Local Similarity 77.8%; Pred. No. 27;
  Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TRLTRKDRG 9
|| ||||:|
Db 94 TRNTRKDQG 102

RESULT 9
A37842
hypothetical protein 1 (xisA 3' region) - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
C;Date: 21-Jun-1991 #sequence_revision 21-Jun-1991 #text_change 09-Jul-2004
C;Accession: A37842
R;Lammers, P.J.; McLaughlin, S.; Papin, S.; Trujillo-Provencio, C.; Ryncarz II, A.J.
J. Bacteriol. 172, 6981-6990, 1990
A;Title: Developmental rearrangement of cyanobacterial nif genes: nucleotide sequence,
A;Reference number: A37842; MUID:91072249; PMID:2123860
A;Accession: A37842
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <LAM>
A;Cross-references: UNIPROT:P29978; GB:M38044
C;Superfamily: Anabaena hypothetical protein 1 (xisA 3' region)

  Query Match 63.6%; Score 35; DB 2; Length 321;
  Best Local Similarity 87.5%; Pred. No. 36;
  Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RLTRKDRG 9
|| |||||
Db 152 RLKRKDRG 159

RESULT 10
AE1987
hypothetical protein alr1448 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AE1987
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE1987
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-384 <KUR>
A;Cross-references: UNIPROT:P29978; GB:BA000019; PIDN:BAB73405.1; PID:g17130795; GSPDB:
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr1448
C;Superfamily: Anabaena hypothetical protein 1 (xisA 3' region)

  Query Match 63.6%; Score 35; DB 2; Length 384;
  Best Local Similarity 87.5%; Pred. No. 43;
  Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RLTRKDRG 9
|| |||||
Db 215 RLKRKDRG 222

```
RESULT 11
C60950
apolipoprotein B-100 - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: C60950
R;Law, A.; Scott, J.
J. Lipid Res. 31, 1109-1120, 1990
A;Title: A cross-species comparison of the apolipoprotein B domain that binds to the LDL
A;Reference number: A60950; MUID:90324804; PMID:2373961
A;Accession: C60950
A;Molecule type: DNA
A;Residues: 1-269 <LAW>
A;Cross-references: UNIPROT:Q60537; UNIPROT:Q60536
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;

  Query Match             62.7%;  Score 34.5;  DB 2;  Length 269;
  Best Local Similarity   81.8%;  Pred. No. 38;
  Matches    9;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              :||||| ||||
Db        216 SRLTRK-RGLK 225


RESULT 12
JH0102
apolipoprotein B - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: JH0102
R;Smith, T.J.
submitted to GenBank, June 1990
A;Reference number: A38864
A;Accession: JH0102
A;Molecule type: DNA
A;Residues: 1-779 <SMI>
A;Cross-references: UNIPROT:Q60536; GB:M35187
A;Note: this is a revision to the sequence from reference JH0101
R;Smith, T.J.; Hautamaa, D.; Maeda, N.
Gene 87, 309-310, 1990
A;Title: Sequence of the putative low-density lipoprotein receptor-binding regions of ap
A;Reference number: JH0101; MUID:90236327; PMID:2332175
A;Contents: annotation
A;Note: this sequence has been revised in reference A38864
C;Genetics:
A;Gene: apoB
C;Superfamily: apolipoprotein B
C;Keywords: atherosclerosis; calcium; cholesterol metabolism; chylomicron; glycoprotein;
F;435-445/Region: receptor binding
F;646-656/Region: receptor binding

  Query Match             62.7%;  Score 34.5;  DB 2;  Length 779;
  Best Local Similarity   81.8%;  Pred. No. 1.1e+02;
  Matches    9;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              :||||| ||||
Db        642 SRLTRK-RGLK 651


RESULT 13
F71164
hypothetical protein PH0515 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: F71164
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71164
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-195 <KAW>
A;Cross-references: UNIPROT:O58251; GB:AP000002; NID:g3236129; PIDN:BAA29603.1; PID:g32
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0515

  Query Match             61.8%;  Score 34;  DB 2;  Length 195;
  Best Local Similarity   60.0%;  Pred. No. 35;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGLK 11
              | : ||||:|
Db        135 RCSEKDRGIK 144


RESULT 14
S73019
daunorubicin resistance protein drrA - Mycobacterium leprae
N;Alternate names: L518_F2_43 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Aug-2004
C;Accession: S73019
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid L518.
A;Reference number: S72591
A;Accession: S73019
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-331 <SMI>
A;Cross-references: UNIPROT:Q49938; EMBL:U00023; NID:g467194; PIDN:AAA17362.1; PID:g467
C;Genetics:
A;Gene: drrA
C;Superfamily: ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;25-216/Domain: ATP-binding cassette homology <ABC>
F;42-49/Region: nucleotide-binding motif A (P-loop)

  Query Match             61.8%;  Score 34;  DB 2;  Length 331;
  Best Local Similarity   87.5%;  Pred. No. 59;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 LTRKDRGL 10
              ||| ||||
Db         58 LTRPDRGL 65


RESULT 15
D64337
16S rRNA 5'-region hypothetical protein 1 homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: D64337
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschi
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: D64337
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-389 <BUL>
A;Cross-references: UNIPROT:Q57747; GB:U67485; GB:L77117; NID:g1591020; PIDN:AAB98286.1
C;Genetics:
A;Map position: REV282927-281758
A;Start codon: GTG
```

C;Superfamily: fructose-1,6-bisphosphatase, archaeal type

Query Match 61.8%; Score 34; DB 2; Length 389;
Best Local Similarity 85.7%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          3 LTRKDRG 9
              :||||||
Db        380 ITRKDRG 386
```

Search completed: January 13, 2005, 01:52:43
Job time : 16.8689 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 13, 2005, 01:15:45 ; Search time 85.8361 Seconds
(without alignments)
73.735 Million cell updates/sec

Title: US-09-823-418-19
Perfect score: 55
Sequence: 1 TRLTRKDRGLK 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	72.7	250	2	Q8DNJ3	Q8dnj3 streptococc
2	40	72.7	2073	1	BIME_EMENI	P24686 emericella
3	39	70.9	438	2	Q99JK2	Q99jk2 mus musculu
4	39	70.9	807	2	Q8Z1J2	Q8z1j2 salmonella
5	38.5	70.0	414	2	Q7YQR5	Q7yqr5 aotus vocif
6	38.5	70.0	596	2	Q28473	Q28473 macaca fasc
7	38.5	70.0	3262	2	Q13788	Q13788 homo sapien
8	38.5	70.0	4563	1	APB_HUMAN	P04114 homo sapien
9	38.5	70.0	4563	2	Q7Z600	Q7z600 homo sapien
10	37	67.3	213	2	Q72PV5	Q72pv5 leptospira
11	37	67.3	213	2	Q8F6D8	Q8f6d8 leptospira
12	37	67.3	213	2	AAS70931	Aas70931 leptospir
13	37	67.3	838	2	Q800W5	Q800w5 brachydanio
14	37	67.3	960	2	Q7N7N1	Q7n7n1 photorhabdu
15	36	65.5	100	1	URE3_CORGL	Q9rhm6 corynebacte
16	36	65.5	100	1	URE3_LACFE	P26931 lactobacill
17	36	65.5	100	2	AAD22478	Aad22478 lactobaci
18	36	65.5	250	2	Q97NY0	Q97ny0 streptococc
19	36	65.5	261	1	EMCN_HUMAN	Q9ulc0 homo sapien
20	36	65.5	283	2	Q7PS51	Q7ps51 anopheles g
21	36	65.5	340	2	Q8GLA1	Q8gla1 streptococc
22	36	65.5	370	2	Q6NDR2	Q6ndr2 rhodopseudo
23	36	65.5	370	2	CAE25486	Cae25486 rhodopseu
24	36	65.5	499	1	GSHR_ARATH	P48641 arabidopsis
25	36	65.5	499	2	AAM98183	Aam98183 arabidops
26	36	65.5	499	2	AAN13086	Aan13086 arabidops
27	36	65.5	499	2	AAP68309	Aap68309 arabidops
28	36	65.5	682	2	Q6NV26	Q6nv26 brachydanio
29	36	65.5	682	2	Q8JHJ0	Q8jhj0 brachydanio
30	36	65.5	682	2	AAH68340	Aah68340 brachydan
31	36	65.5	1007	2	Q8RWG0	Q8rwg0 arabidopsis
32	35	63.6	78	2	Q9JZ39	Q9jz39 neisseria m
33	35	63.6	111	2	Q95X71	Q95x71 caenorhabdi
34	35	63.6	198	2	Q8SUL5	Q8sul5 encephalito
35	35	63.6	220	2	Q8DIV0	Q8div0 synechococc
36	35	63.6	236	2	Q8YMB7	Q8ymb7 anabaena sp
37	35	63.6	361	2	Q840G9	Q840g9 actinobacil
38	35	63.6	374	2	Q8KIP9	Q8kip9 pseudomonas
39	35	63.6	374	2	Q8KIV1	Q8kiv1 pseudomonas
40	35	63.6	384	1	YE48_ANASP	P29978 anabaena sp
41	35	63.6	391	2	Q8FLE8	Q8fle8 corynebacte
42	35	63.6	480	2	Q9H834	Q9h834 homo sapien
43	35	63.6	506	2	Q6Z7C0	Q6z7c0 oryza sativ
44	35	63.6	506	2	BAD07650	Bad07650 oryza sat
45	35	63.6	506	2	BAD07928	Bad07928 oryza sat

ALIGNMENTS

RESULT 1
Q8DNJ3
ID Q8DNJ3 PRELIMINARY; PRT; 250 AA.
AC Q8DNJ3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-ferric iron transport.
GN Name=fecE; OrderedLocusNames=spr1686;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AE008534; AAL00489.1; -.
DR PIR; D98082; D98082.
DR HSSP; P06611; 1L7V.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 250 AA; 28528 MW; FFC841A29F30033D CRC64;

Query Match 72.7%; Score 40; DB 2; Length 250;
Best Local Similarity 70.0%; Pred. No. 7.2;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TRLTRKDRGL 10
              :|||:||||:
Db         47 SRLTKKDRGV 56
```

```
RESULT 2
BIME_EMENI
ID   BIME_EMENI     STANDARD;      PRT;  2073 AA.
AC   P24686;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Negative regulator of mitosis.
GN   Name=BIME;
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90375468; PubMed=1697851;
RA   Engle D.B., Osmani S.A., Osmani A.H., Rosborough S., Xiang X.,
RA   Morris N.R.;
RT   "A negative regulator of mitosis in Aspergillus is a putative
RT   membrane-spanning protein.";
RL   J. Biol. Chem. 265:16132-16137(1990).
CC   -!- FUNCTION: Negative regulator of mitosis in E.nidulans. This
CC       protein is part of a regulatory pathway that includes the nimA
CC       protein kinase. It is required to prevent premature entry into
CC       mitosis. Mutations to this protein both cause cells to enter
CC       mitosis and prevent them from leaving mitosis.
CC   -!- SIMILARITY: Contains 4 PC repeats.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M59705; AAA51478.1; -.
DR   PIR; A37879; BWASBE.
DR   InterPro; IPR002015; APC_proteasome.
DR   Pfam; PF01851; PC_rep; 4.
KW   Mitosis; Repeat; Transmembrane.
FT   DOMAIN      342    353       Nuclear localization signal (Potential).
FT   TRANSMEM   1623   1643       Potential.
FT   TRANSMEM   1685   1703       Potential.
FT   TRANSMEM   1746   1764       Potential.
SQ   SEQUENCE   2073 AA;  229178 MW;  05E4E81EADDF51E4 CRC64;

  Query Match             72.7%;  Score 40;  DB 1;  Length 2073;
  Best Local Similarity   80.0%;  Pred. No. 88;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDRGL 10
              |||||: |||
Db        832 TRLTRRKRGL 841


RESULT 3
Q99JK2
ID   Q99JK2      PRELIMINARY;      PRT;   438 AA.
AC   Q99JK2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Lars protein.
GN   Name=Lars;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Czech II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Czech II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006060; AAH06060.1; -.
DR   MGD; MGI:1913808; Lars.
DR   InterPro; IPR002088; PPTA.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   PROSITE; PS00904; PPTA; UNKNOWN_1.
SQ   SEQUENCE   438 AA;  49840 MW;  2E730A99D65AFBF0 CRC64;

  Query Match             70.9%;  Score 39;  DB 2;  Length 438;
  Best Local Similarity   70.0%;  Pred. No. 23;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGLK 11
              |||: |||:|
Db        361 RLTKTDRGIK 370


RESULT 4
Q8Z1J2
ID   Q8Z1J2      PRELIMINARY;      PRT;   807 AA.
AC   Q8Z1J2; Q7C5G1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein STY4573.
GN   OrderedLocusNames=STY4573, t4270;
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
DR   EMBL; AL627282; CAD09348.1; -.
DR   EMBL; AE016848; AAO71728.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   807 AA;  90698 MW;  577C1DA31611BCF0 CRC64;

  Query Match             70.9%;  Score 39;  DB 2;  Length 807;
  Best Local Similarity   88.9%;  Pred. No. 47;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              |||| ||||
Db        508 RLTRADRGL 516


RESULT 5
Q7YQR5
ID   Q7YQR5      PRELIMINARY;      PRT;   414 AA.
AC   Q7YQR5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Apolipoprotein B 100 (Fragment).
GN   Name=apoB-100;
OS   Aotus vociferans (Spix's owl monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX   NCBI_TaxID=57176;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22761261; PubMed=12878460;
RA   Amrine-Madsen H., Koepfli K.-P., Wayne R.K., Springer M.S.;
RT   "A new phylogenetic marker, apolipoprotein B, provides compelling
RT   evidence for eutherian relationships.";
RL   Mol. Phylogenet. Evol. 28:225-240(2003).
DR   EMBL; AF548396; AAP97352.1; -.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     414    414
SQ   SEQUENCE   414 AA;  45955 MW;  EEFA8492157E1BDE CRC64;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 414;
  Best Local Similarity   90.9%;  Pred. No. 27;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db        258 TRLTRK-RGLK 267


RESULT 6
Q28473
ID   Q28473      PRELIMINARY;      PRT;   596 AA.
AC   Q28473;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Apolipoprotein B (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92075708; PubMed=1742325;
RA   Pape M.E., Castle C.K., Murray R.W., Funk G.M., Hunt C.E.,
RA   Marotti K.R., Melchior G.W.;
RT   "Apo B metabolism in the cynomolgus monkey: evidence for post-
RT   transcriptional regulation.";
RL   Biochim. Biophys. Acta 1086:326-334(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Murray R.;
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X15737; CAA33755.1; -.
DR   PIR; S32802; S32802.
KW   Lipoprotein.
FT   NON_TER       1      1
FT   NON_TER     596    596
SQ   SEQUENCE   596 AA;  66757 MW;  B13BBA74E25C3120 CRC64;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 596;
  Best Local Similarity   90.9%;  Pred. No. 42;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db        226 TRLTRK-RGLK 235


RESULT 7
Q13788
ID   Q13788      PRELIMINARY;      PRT;  3262 AA.
AC   Q13788;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   APOB protein (Fragment).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87191999; PubMed=2883086;
RA   Carlsson P., Darnfors C., Olofsson S.O., Bjursell G.;
RT   "Analysis of the human apolipoprotein B gene; complete structure of
RT   the B-74 region.";
RL   Gene 49:29-51(1986).
DR   EMBL; M15421; AAA51758.1; -.
DR   PIR; A27850; LPHUB.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0005319; F:lipid transporter activity; NAS.
DR   GO; GO:0006869; P:lipid transport; NAS.
FT   NON_TER       1      1
SQ   SEQUENCE   3262 AA;  370140 MW;  56603BC0618DD40D CRC64;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 3262;
  Best Local Similarity   90.9%;  Pred. No. 3.2e+02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db       2084 TRLTRK-RGLK 2093


RESULT 8
APB_HUMAN
ID   APB_HUMAN      STANDARD;      PRT;  4563 AA.
AC   P04114; O00502; Q13787;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein
```

DE B-48 (Apo B-48)].
GN Name=APOB;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87016385; PubMed=3763409;
RA Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
RA Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
RT "Complete cDNA and derived protein sequence of human apolipoprotein B-
RT 100.";
RL Nucleic Acids Res. 14:7501-7503(1986).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT GLU-4181.
RX MEDLINE=88003974; PubMed=3652907;
RA Ludwig E.H., Blackhart B.D., Pierotti V.R., Caiati L., Fortier C.,
RA Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
RT "DNA sequence of the human apolipoprotein B gene.";
RL DNA 6:363-372(1987).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS ILE-98 AND GLU-4181.
RX MEDLINE=87008488; PubMed=3759943;
RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
RA Gotto A.M. Jr., Chan L.;
RT "The complete cDNA and amino acid sequence of human apolipoprotein B-
RT 100.";
RL J. Biol. Chem. 261:12918-12921(1986).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=87041416; PubMed=3464946;
RA Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
RA Lee N., Brewer H.B. Jr.;
RT "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
RT derived amino acid sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=87161758; PubMed=3030729;
RA Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D.,
RA Zannis V.I.;
RT "The complete sequence and structural analysis of human apolipoprotein
RT B-100: relationship between apoB-100 and apoB-48 forms.";
RL EMBO J. 5:3495-3507(1986).
RN [6]
RP SEQUENCE OF 709-906 FROM N.A.
RX MEDLINE=85270450; PubMed=3860836;
RA Deeb S.S., Motulsky A.G., Albers J.J.;
RT "A partial cDNA clone for human apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
RN [7]
RP SEQUENCE OF 3056-3159 FROM N.A.
RX MEDLINE=86041888; PubMed=3903660;
RA Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
RA Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
RT "Human apolipoprotein B: identification of cDNA clones and
RT characterization of mRNA.";
RL Nucleic Acids Res. 13:6937-6953(1985).
RN [8]
RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A., AND VARIANT GLU-4181.
RX MEDLINE=86093680; PubMed=3841204;
RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
RA Bjursell G.;
RT "Molecular cloning of human apolipoprotein B cDNA.";
RL Nucleic Acids Res. 13:8813-8826(1985).
RN [9]
RP SEQUENCE OF 3109-4563 FROM N.A.
RX MEDLINE=85300528; PubMed=2994225;
RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Urdea M.S.,
RA Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M.,
RA Priestley L.M., Robertson E., Rall L.B., Betsholtz C., Shows T.B.,
RA Mahley R.W., Scott J.;
RT "Human apolipoprotein B: structure of carboxyl-terminal domains, sites
RT of gene expression, and chromosomal localization.";
RL Science 230:37-43(1985).
RN [10]
RP SEQUENCE OF 1-291 FROM N.A.
RX MEDLINE=86149325; PubMed=3513177;
RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V.,
RA Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;
RT "Isolation of a cDNA clone encoding the amino-terminal region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
RN [11]
RP SEQUENCE OF 1-1670 FROM N.A., AND VARIANT ILE-98.
RX MEDLINE=86287319; PubMed=3461454;
RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanaka M.,
RA Hort Y.J., Hjerrild K.A., Chen G.C., Kane J.P.;
RT "Analysis of cDNA clones encoding the entire B-26 region of human
RT apolipoprotein B.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
RN [12]
RP PARTIAL SEQUENCE, AND IDENTIFICATION OF APO-B48.
RX MEDLINE=88018019; PubMed=3659919;
RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H.,
RA Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
RA Gotto A.M. Jr., Li W.-H., Chan L.;
RT "Apolipoprotein B-48 is the product of a messenger RNA with an organ-
RT specific in-frame stop codon.";
RL Science 238:363-366(1987).
RN [13]
RP DOMAINS.
RX MEDLINE=87039351; PubMed=3773997;
RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr.,
RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
RA Levy-Wilson B., Scott J.;
RT "Complete protein sequence and identification of structural domains of
RT human apolipoprotein B.";
RL Nature 323:734-738(1986).
RN [14]
RP DOMAINS.
RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
RA Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
RT "Sequence, structure, receptor-binding domains and internal repeats of
RT human apolipoprotein B-100.";
RL Nature 323:738-742(1986).
RN [15]
RP CALCIUM-BINDING DATA.
RX MEDLINE=86242245; PubMed=3087360;
RA Dashti N., Lee D.M., Mok T.;
RT "Apolipoprotein B is a calcium binding protein.";
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
RN [16]
RP PALMITOYLATION OF CYS-1112.
RX MEDLINE=20143590; PubMed=10679026;
RA Zhao Y., McCabe J.B., Vance J., Berthiaume L.G.;
RT "Palmitoylation of apolipoprotein B is required for proper
RT intracellular sorting and transport of cholesteroyl esters and
RT triglycerides.";
RL Mol. Biol. Cell 11:721-734(2000).
RN [17]
RP VARIANT SER-4338.
RX MEDLINE=91071750; PubMed=1979313;
RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
RA Cunny G., Cambien F., Roizes G.;
RT "Detection by denaturing gradient gel electrophoresis of a new
RT polymorphism in the apolipoprotein B gene.";
RL Hum. Genet. 86:91-93(1990).
RN [18]
RP VARIANT FDB GLN-3527.
RX MEDLINE=89098975; PubMed=2563166;
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
RA McCarthy B.J.;

```
RT   "Association between a specific apolipoprotein B mutation and familial
RT   defective apolipoprotein B-100.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
RN   [19]
RP   VARIANT LEU-2739.
RX   MEDLINE=91016974; PubMed=2216805;
RA   Huang L.-S., Gavish D., Breslow J.L.;
RT   "Sequence polymorphism in the human apoB gene at position 8344.";
RL   Nucleic Acids Res. 18:5922-5922(1990).
RN   [20]
RP   VARIANT FDB CYS-3558.
RX   MEDLINE=95190020; PubMed=7883971;
RA   Pullinger C.R., Hennessy L.K., Chatterton J.E., Liu W., Love J.A.,
RA   Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
RT   "Familial ligand-defective apolipoprotein B. Identification of a new
RT   mutation that decreases LDL receptor binding affinity.";
RL   J. Clin. Invest. 95:1225-1234(1995).
RN   [21]
RP   VARIANTS LEU-1437; SER-1914; LYS-2566; THR-3121; ALA-3945; MET-4128
RP   AND THR-4481.
RX   MEDLINE=97044521; PubMed=8889592;
RA   Poirier O., Ricard S., Behague I., Souriau C., Evans A.E.,
RA   Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
RT   "Detection of new variants in the apolipoprotein B (Apo B) gene by
RT   PCR-SSCP.";
RL   Hum. Mutat. 8:282-285(1996).
RN   [22]
RP   VARIANTS FDB GLN-3527 AND CYS-3558.
RX   MEDLINE=97403938; PubMed=9259199;
RA   Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
RA   Krempf M., Giraudet P., Junien C., Boileau C.;
RT   "Familial ligand-defective apolipoprotein B-100: simultaneous
RT   detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a French
RT   population.";
RL   Hum. Mutat. 10:160-163(1997).
RN   [23]
RP   VARIANTS SER-1914; ARG-1923; LEU-2739; ASP-3319; THR-3427; GLN-3432
RP   AND ILE-3921.
RX   MEDLINE=98141125; PubMed=9490296;
RA   Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
RT   "Screening for mutations of the apolipoprotein B gene causing
RT   hypocholesterolemia.";
RL   Hum. Genet. 102:44-49(1998).
CC   -!- FUNCTION: Apolipoprotein B is a major protein constituent of
CC       chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo
CC       B-100 functions as a recognition signal for the cellular binding
CC       and internalization of LDL particles by the apoB/E receptor.
CC   -!- SUBCELLULAR LOCATION: Secreted.

  Query Match             70.0%;  Score 38.5;  DB 1;  Length 4563;
  Best Local Similarity   90.9%;  Pred. No. 4.7e+02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db       3385 TRLTRK-RGLK 3394


RESULT 9
Q7Z600
ID   Q7Z600      PRELIMINARY;      PRT;  4563 AA.
AC   Q7Z600;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Apolipoprotein B (Including Ag(X) antigen).
GN   Name=APOB;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., da Ponte S.H., Hastings N.C.,
RA   Ahearn M.O., Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY324608; AAP72970.1; -.
DR   GO; GO:0005319; F:lipid transporter activity; IEA.
DR   GO; GO:0006869; P:lipid transport; IEA.
DR   InterPro; IPR009454; DUF1081.
DR   InterPro; IPR001747; Lipid_transprt_N.
DR   Pfam; PF06448; DUF1081; 1.
DR   Pfam; PF01347; Vitellogenin_N; 1.
DR   SMART; SM00638; LPD_N; 1.
KW   Lipoprotein.
SQ   SEQUENCE   4563 AA;  515553 MW;  030B34167CEDC63C CRC64;

  Query Match             70.0%;  Score 38.5;  DB 2;  Length 4563;
  Best Local Similarity   90.9%;  Pred. No. 4.7e+02;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy          1 TRLTRKDRGLK 11
              |||||| ||||
Db       3385 TRLTRK-RGLK 3394


RESULT 10
Q72PV5
ID   Q72PV5      PRELIMINARY;      PRT;  213 AA.
AC   Q72PV5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   3-methyladenine DNA glycosylase.
GN   Name=alkA; OrderedLocusNames=LIC12362;
OS   Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar
OS   Copenhageni).
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=44275;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Fiocruz L1-130;
RX   PubMed=15028702;
RA   Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
RA   Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,
RA   Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H.,
RA   Coutinho L.L., Degrave W.M., Dellagostin O.A., El-Dorry H.,
RA   Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Giglioti E.A.,
RA   Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
RA   Jeronimo S.M.B., Junqueira-de-Azevedo I.L.M., Kimura E.T.,
RA   Kuramae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,
RA   de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A.,
RA   Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
RA   Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Sluys M.A.;
RT   "Comparative genomics of two Leptospira interrogans serovars reveals
RT   novel insights into physiology and pathogenesis.";
RL   J. Bacteriol. 186:2164-2172(2004).
DR   EMBL; AE017296; AAS70931.1; -.
DR   InterPro; IPR011257; DNA_glycsylse.
DR   InterPro; IPR003265; Endo_3c.
DR   Pfam; PF00730; HhH-GPD; 1.
DR   SMART; SM00478; ENDO3c; 1.
KW   Complete proteome.
SQ   SEQUENCE   213 AA;  24639 MW;  73E6CA3E0C737487 CRC64;

  Query Match             67.3%;  Score 37;  DB 2;  Length 213;
  Best Local Similarity   77.8%;  Pred. No. 26;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              :|:||||||
Db         19 QLSRKDRGL 27


RESULT 11
```

```
Q8F6D8
ID   Q8F6D8      PRELIMINARY;      PRT;   213 AA.
AC   Q8F6D8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   DNA-3-methyladenine glycosylase (EC 3.2.2.21).
GN   Name=magl; OrderedLocusNames=LA1370;
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX   MEDLINE=22598143; PubMed=12712204;
RA   Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA   Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA   Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA   Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA   Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA   Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,
RA   Xu J.-G., Zhao G.-P.;
RT   "Unique physiological and pathogenic features of Leptospira
RT   interrogans revealed by whole-genome sequencing.";
RL   Nature 422:888-893(2003).
DR   EMBL; AE011316; AAN48569.1; -.
DR   GO; GO:0003905; F:alkylbase DNA N-glycosylase activity; IEA.
DR   GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR   GO; GO:0006284; P:base-excision repair; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR011257; DNA_glycsylse.
DR   InterPro; IPR003265; Endo_3c.
DR   Pfam; PF00730; HhH-GPD; 1.
DR   SMART; SM00478; ENDO3c; 1.
KW   Complete proteome; Glycosidase; Hydrolase.
SQ   SEQUENCE   213 AA;  24567 MW;  01C4CA3DB9259732 CRC64;

  Query Match             67.3%;  Score 37;  DB 2;  Length 213;
  Best Local Similarity   77.8%;  Pred. No. 26;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              :|:||||||
Db         19 QLSRKDRGL 27


RESULT 12
AAS70931
ID   AAS70931    PRELIMINARY;      PRT;   213 AA.
AC   AAS70931;
DT   24-MAR-2004 (TrEMBLrel. 27, Created)
DT   24-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   11-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE   3-methyladenine DNA glycosylase.
GN   ALKA OR LIC12362.
OS   Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar
OS   Copenhageni).
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=44275;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Fiocruz L1-130;
RX   PubMed=15028702;
RA   Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
RA   Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,
RA   Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H.,
RA   Coutinho L.L., Degrave W.M., Dellagostin O.A., El-Dorry H.,
RA   Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Giglioti E.A.,
RA   Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
RA   Jeronimo S.M.B., Junqueira-de-Azevedo I.L.M., Kimura E.T.,
RA   Kuramae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,
RA   de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A.,
RA   Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
RA   Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Sluys M.A.;
RT   "Comparative genomics of two Leptospira interrogans serovars reveals
RT   novel insights into physiology and pathogenesis.";
RL   J. Bacteriol. 186:2164-2172(2004).
DR   EMBL; AE017296; AAS70931.1; -.
SQ   SEQUENCE   213 AA;  24639 MW;  73E6CA3E0C737487 CRC64;

  Query Match             67.3%;  Score 37;  DB 2;  Length 213;
  Best Local Similarity   77.8%;  Pred. No. 26;
  Matches    7;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              :|:||||||
Db         19 QLSRKDRGL 27


RESULT 13
Q800W5
ID   Q800W5      PRELIMINARY;      PRT;   838 AA.
AC   Q800W5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Frodo2.
GN   Name=frd2;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21984492; PubMed=11941372;
RA   Gloy J., Hikasa H., Sokol S.Y.;
RT   "Frodo interacts with Dishevelled to transduce Wnt signals.";
RL   Nat. Cell Biol. 4:351-357(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   PubMed=15188426;
RA   Gillhouse M., Wagner Nyholm M., Hikasa H., Sokol S.Y., Grinblat Y.;
RT   "Two Frodo/Dapper homologs are expressed in the developing brain and
RT   mesoderm of zebrafish.";
RL   Dev. Dyn. 230:403-409(2004).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hikasa H., Sokol S.Y.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY208969; AAO49711.2; -.
DR   ZFIN; ZDB-GENE-030131-9975; dact.
SQ   SEQUENCE   838 AA;  91349 MW;  6906B5ECC35012BC CRC64;

  Query Match             67.3%;  Score 37;  DB 2;  Length 838;
  Best Local Similarity   72.7%;  Pred. No. 1.3e+02;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TRLTRKDRGLK 11
              ||| |:| |||
Db         86 TRLRRQDVGLK 96


RESULT 14
Q7N7N1
ID   Q7N7N1      PRELIMINARY;      PRT;   960 AA.
AC   Q7N7N1;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Similar to unknown protein.
GN   OrderedLocusNames=plu1085;
OS   Photorhabdus luminescens (subsp. laumondii).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Photorhabdus.
OX   NCBI_TaxID=141679;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TT01;
RX   MEDLINE=22957627; PubMed=14528314;
RA   Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA   Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA   Dassa E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,
RA   Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA   Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT   "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT   luminescens.";
RL   Nat. Biotechnol. 21:1307-1313(2003).
DR   EMBL; BX571862; CAE13380.1; -.
DR   PhotoList; plu1085; -.
KW   Complete proteome.
SQ   SEQUENCE   960 AA;  107431 MW;  1329711136B972B1 CRC64;

  Query Match             67.3%;  Score 37;  DB 2;  Length 960;
  Best Local Similarity   77.8%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 RLTRKDRGL 10
              |||| |||:
Db        646 RLTRADRGM 654


RESULT 15
URE3_CORGL
ID   URE3_CORGL     STANDARD;      PRT;   100 AA.
AC   Q9RHM6;
DT   29-MAR-2004 (Rel. 43, Created)
DT   29-MAR-2004 (Rel. 43, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase gamma subunit).
GN   Name=ureA; OrderedLocusNames=Cgl0084, cg0113;
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13869;
RA   Puskas L.G., Inui M., Yukawa H.;
RT   "Structure and transcriptional regulation of the urease operon of
RT   Corynebacterium glutamicum.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nolden L., Beckers G., Moeckel B., Nampoothiri M., Pfefferle W.,
RA   Kraemer R., Burkovski A.;
RT   "Urease of Corynebacterium glutamicum: sequence and organisation of
RT   corresponding genes and investigation of activity.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RX   PubMed=12948626; DOI=10.1016/S0168-1656(03)00154-8;
RA   Kalinowski J., Bathe B., Bartels D., Bischoff N., Bott M.,
RA   Burkovski A., Dusch N., Eggeling L., Eikmanns B.J., Gaigalat L.,
RA   Goesmann A., Hartmann M., Huthmacher K., Kraemer R., Linke B.,
RA   McHardy A.C., Meyer F., Moeckel B., Pfefferle W., Puehler A.,
RA   Rey D.A., Rueckert C., Rupp O., Sahm H., Wendisch V.F., Wiegraebe I.,
RA   Tauch A.;
RT   "The complete Corynebacterium glutamicum ATCC 13032 genome sequence
RT   and its impact on the production of L-aspartate-derived amino acids
RT   and vitamins.";
RL   J. Biotechnol. 104:5-25(2003).
CC   -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC   -!- SUBUNIT: (Alpha, beta, gamma)(3) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the urease gamma subunit family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB029154; BAA88552.1; -.
DR   EMBL; AJ251883; CAB81935.1; -.
DR   EMBL; AP005274; BAB97477.1; -.
DR   EMBL; BX927148; CAF18652.1; -.
DR   HSSP; P41022; 1UBP.
DR   HAMAP; MF_00739; -; 1.
DR   InterPro; IPR002026; Urease_gamma.
DR   Pfam; PF00547; Urease_gamma; 1.
DR   ProDom; PD002319; Urease_gamma; 1.
DR   TIGRFAMs; TIGR00193; urease_gam; 1.
KW   Complete proteome; Hydrolase.
SQ   SEQUENCE   100 AA;  11245 MW;  A48F4DC0EABD9567 CRC64;

  Query Match             65.5%;  Score 36;  DB 1;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 RKDRGLK 11
              |||||||
Db         23 RKDRGLK 29


Search completed: January 13, 2005, 01:51:10
Job time : 87.8361 secs
```

THIS PAGE LEFT BLANK

THIS PAGE IS BLANK